

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2005, 18:09:49 ; Search time 1816 Seconds  
(without alignments)  
480.283 Million cell updates/sec

Title: US-10-677-982-1  
Perfect score: 18  
Sequence: 1 GTTCTACATATGCGCG 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapect 1.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb Da:\*  
2: gb Hcg:\*  
3: gb In:\*  
4: gb Om:\*  
5: gb Ov:\*  
6: gb Pat:\*  
7: gb Ph:\*  
8: gb Pl:\*  
9: gb Pr:\*  
10: gb Ro:\*  
11: gb Ste:\*  
12: gb Sy:\*  
13: gb Un:\*  
14: gb Vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	18	6	ARS34202 Sequence
2	18	100.0	18	6	ARS34203 Sequence
3	18	100.0	173185	10	AC125317 Mus muscu
4	18	100.0	190369	10	AC119908 Mus muscu
5	16.4	91.1	677	8	AK109255 Oryza sat
6	16.4	91.1	47636	2	AC014385 Drosophi1
7	16.4	91.1	92958	2	DMBR1J17 AL121811 Drosophi1
8	16.4	91.1	144233	8	AP003240 Oryza sat
9	16.4	91.1	165928	3	AC023725 Drosophi1
10	16.4	91.1	172784	3	AC105352 Drosophi1
11	16.4	91.1	181831	3	AC107403 Drosophi1
12	16.4	91.1	309037	3	AE003429 Drosophi1
13	16.4	91.1	321355	2	AC128492 Rattus no
14	16.4	88.9	303750	1	AB016931 Bacteroid
15	15.4	85.6	549	8	AK063186 Oryza sat
16	15.4	85.6	855	8	AK066807 Oryza sat
17	15.4	85.6	1519	6	CQ593598 Sequence
18	15.4	85.6	2217	3	AY052136 Drosophi1
19	15.4	85.6	2336	3	BT001590 Drosophi1

20	15.4	85.6	2998	5	BC079943	Xenopus t
21	15.4	85.6	4927	6	CQ593597	Sequence
22	15.4	85.6	9847	2	AC020439	Drosophi1
23	15.4	85.6	10029	1	AE011108	Methanosa
24	15.4	85.6	119910	8	AC134238	Oryza sat
25	15.4	85.6	132637	2	AC006092	Drosophi1
26	15.4	85.6	157481	2	CR753894	Mus muscu
27	15.4	85.6	167088	2	AC142365	Rattus no
28	15.4	85.6	177339	3	AC008231	Drosophi1
29	15.4	85.6	185184	10	AL627411	Mouse DNA
30	15.4	85.6	187399	9	AC067942	Homo sapi
31	15.4	85.6	188633	3	AC007175	Drosophi1
32	15.4	85.6	199520	2	AC110420	Homo sapi
33	15.4	85.6	200050	1	AL591973	Listeria
34	15.4	85.6	217306	2	CR354548	Mus muscu
35	15.4	85.6	225957	2	AC118075	Rattus no
36	15.4	85.6	253169	2	AC094461	Rattus no
37	15.4	85.6	301450	1	AP003188	Clostridi
38	15.4	85.6	302101	1	AE016784	Pseudomon
39	15.4	85.6	305150	3	AE003453	Drosophi1
40	15.4	85.6	317136	2	AC098150	Rattus no
41	15.4	85.6	349980	6	AX638811	Sequence
42	15	83.3	148641	9	HS111909	Human DNA
43	15	83.3	158122	5	AL935169	Zebrafish
44	15	83.3	183225	2	BX323836	Danio rer
45	15	83.3	197484	2	CR391947	Danio rer

#### ALIGNMENTS

RESULT 1  
LOCUS ARS34202  
DEFINITION Sequence 1 from patent US 6733971.  
ACCESSION ARS34202  
VERSION ARS34202.1 GI:53924250  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE  
AUTHORS 1 (bases 1 to 18)  
TITLE Beachy,P.A., Tsai,M.-J., Tsai,S., Krishnan,V. and Chen,C.-H.  
METHOD for identifying an agent that affects a hedgehog signaling pathway  
JOURNAL Patent: US 6733971-A 1 11-MAY-2004;  
FEATURES  
source Location/Qualifiers  
1..18  
/organism="unknown"  
/mol\_type="genomic DNA"

#### ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 18;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GTTCTACATATGCGCG 18  
Db 1 GTTCTACATATGCGCG 18

#### RESULT 2

LOCUS ARS34203/c  
DEFINITION Sequence 2 from patent US 6733971.  
ACCESSION ARS34203  
VERSION ARS34203.1 GI:53924251  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 18)  
AUTHORS Beachy,P.A., Tsai,M.-J., Tsai,S., Krishnan,V. and Chen,C.-H.



```

repeat_region      12555. .12729
/rpc_family="B2"
repeat_region      12794. .13038
/rpc_family="MERL_type"
repeat_region      13551. .13739
/rpc_family="B4"
repeat_region      13744. .14135
/rpc_family="L1"
repeat_region      14169. .14359
/rpc_family="B2"
repeat_region      14395. .14711
/rpc_family="ERVK"
repeat_region      14753. .14835
/rpc_family="ERV1"
repeat_region      14836. .15263
/rpc_family="L1"
repeat_region      15313. .15392
/rpc_family="ID"
repeat_region      15404. .15594
/rpc_family="MIR"
repeat_region      17002. .17241
/rpc_family="L1"
repeat_region      17309. .17704
/rpc_family="L1"
repeat_region      17705. .17885
/rpc_family="L1"
repeat_region      18241. .18487
/rpc_family="B4"
repeat_region      18427. .18499
/rpc_family="B2"
repeat_region      18549. .18814
/rpc_family="L1"
repeat_region      18872. .18983
/rpc_family="L1"
repeat_region      19014. .19501
/rpc_family="L1"
repeat_region      21526. .21724
/rpc_family="B2"
repeat_region      22573. .22982
/rpc_family="MALR"
repeat_region      23290. .23587
/rpc_family="B4"
repeat_region      25375. .25517
/rpc_family="B4"
repeat_region      26791. .26994
/rpc_family="B2"
repeat_region      27298. .27399
/rpc_family="L1"
repeat_region      27635. .27811
/rpc_family="B2"
repeat_region      28131. .28332
/rpc_family="MALR"
repeat_region      28378. .28483
/rpc_family="MALR"
repeat_region      28910. .29063
/rpc_family="B4"
repeat_region      31816. .31952
/rpc_family="Alu"
repeat_region      32142. .32250
/rpc_family="L1"
repeat_region      32685. .32994
/rpc_family="MALR"
repeat_region      33298. .33370
/rpc_family="L1"
repeat_region      33371. .34354
/rpc_family="L1"
repeat_region      34355. .34840
/rpc_family="L1"
repeat_region      34976. .35171
/rpc_family="B2"
repeat_region      36213. .37443
/rpc_family="L1"
repeat_region      37526. .37669

```

```

/rpc_family="Alu"
Query Match      100.0%; Score 18; DB 10; Length 173185;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GTTCTACATATATGCGCG 18
Db      163463 GTTCTACATATATGCGCG 163446

RESULT 4
AC119908
LOCUS
DEFINITION
AC119908
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

/rpc_family="Alu"
AC119908      190369 bp      DNA      linear      ROD 08-SEP-2004
Mus musculus chromosome 7, clone RP24-24064, complete sequence.
AC119908
AC119908.16      GI:51921399
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 190369)
Mus musculus chromosome 7, clone RP24-24064
Unpublished
2 (bases 1 to 190369)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Borkhagalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Johnson,R., Jones,C.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Labocque,K., Lamazares,R.,
Landers,T., Lenoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
McCleary,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Plunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 190369)
Birren,B., Nusbaum,C., Lander,E., Aboueleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Borkhagalter,B., Camarata,J., Chang,J.,
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Court,B.,
Dearellano,K., Diaz,J.S., Dodge,S., Doyle,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
McCleary,C., MacDonald,P., Major,J., Manning,U., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Plunkhang,P., Pierre,N., Rachupka,A., Ramsamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,

```

TITLE  
Direct Submission  
JOURNAL  
Submitted (30-JUL-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE  
4 (bases 1 to 190369)  
AUTHORS  
Birren, B., Nusbaum, C., Lander, E., Abouelella, A., Allen, N.,

Zimmer, A. and Zody, M.  
Anderson, M., Anderson, S., Arachchi, H. M., Barua, N., Bastien, V.,  
Blom, T., Boguslavsky, L., Bouckhalter, B., Camarata, J., Chang, J.,  
Chapel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B.,  
DeRisi, J., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L.,  
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,  
Galagan, J., Gardner, S., Graham, L., Grand-Pierre, N., Hafez, I.,  
Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karas, A., Kelle, C., Landers, T.,  
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,  
McCarthy, M., Meldrum, J., Menus, L., Milnova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,  
Reita, R., Rie, C., Rogov, P., Roman, J., Schauer, S., Schuback, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,  
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,  
Zimmer, A. and Zody, M.  
Submitted (08-SEP-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT  
On Sep 8, 2004 this sequence version replaced gi:50839014.  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/XM/RepeatMasker.html  
Smt, A.R.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/XM/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@road.mit.edu  
----- Project Information  
Center project name: L25288  
Center clone name: 240\_G\_4  
-----

FEATURES  
source  
Some of the sequence contained within base pairs 150144 to the end  
of the clone was stolen from accession AC125317.  
location/Qualifiers  
1..190369

/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="7"  
/map="7"  
/clone="RP24-240G4"  
1..4  
/clone\_id="RP24-240G4"  
misc\_feature  
1..4  
/note="clone boundary  
clone end:SP6  
site:MboI"  
repeat\_region  
376..503  
/rpt\_family="B1P"  
repeat\_region  
514..534  
/rpt\_family="(CAAAA)n"  
repeat\_region  
1553..1603  
/rpt\_family="(CA)n"  
repeat\_region  
2799..2836  
/rpt\_family="(GGA)n"  
repeat\_region  
3129..3271  
/rpt\_family="B1\_MM"  
repeat\_region  
3275..3327  
/rpt\_family="(TGA)n"  
repeat\_region  
3435..3630  
/rpt\_family="Tigres5"  
repeat\_region  
3860..3943

repeat\_region  
/rpt\_family="PB1"  
3944..3989  
/rpt\_family="ID\_B1"  
repeat\_region  
4002..4640  
/rpt\_family="(GGA)n"  
4500  
/note="unresolved VNTR: <500 bp of sequence missing here"  
repeat\_region  
4641..4819  
/rpt\_family="(GAA)n"  
repeat\_region  
4873..4913  
/rpt\_family="(CAAA)n"  
repeat\_region  
5263..5550  
/rpt\_family="B4"  
repeat\_region  
5910..6002  
/rpt\_family="(TCTG)n"  
repeat\_region  
6015..6055  
/rpt\_family="(CA)n"  
repeat\_region  
10679..10749  
/rpt\_family="RSINB1"  
repeat\_region  
10823..10972  
/rpt\_family="(TGA)n"  
repeat\_region  
11070..11351  
/rpt\_family="MERSB8"  
repeat\_region  
11399..11470  
/rpt\_family="(TGC)n"  
repeat\_region  
complement(11516)..11643  
/rpt\_family="B1\_MM"  
repeat\_region  
complement(11649)..11910  
/rpt\_family="B4"  
repeat\_region  
12246..12418  
/rpt\_family="RSINB1"  
repeat\_region  
12937..13142  
/rpt\_family="B3"  
repeat\_region  
complement(13343)..13427  
/rpt\_family="PBID9"  
repeat\_region  
15533..15581  
/rpt\_family="MIR"  
repeat\_region  
18486..18512  
/rpt\_family="(A)n"  
repeat\_region  
18805..18984  
/rpt\_family="(GAAA)n"  
repeat\_region  
18985..19037  
/rpt\_family="A-rich"  
repeat\_region  
19454..19569  
/rpt\_family="PBID10"  
repeat\_region  
19535..19589  
/note="single clone coverage"  
repeat\_region  
19570..19632  
/rpt\_family="(GAA)n"  
repeat\_region  
19633..19810  
/rpt\_family="(GAGA)n"  
repeat\_region  
complement(20553)..20708  
/rpt\_family="ID\_B1"  
repeat\_region  
20990..21035  
/rpt\_family="(TGA)n"  
repeat\_region  
complement(21178)..21286  
/rpt\_family="PBID7"  
repeat\_region  
21370..21417  
/rpt\_family="(TA)n"  
repeat\_region  
22368..22385  
/rpt\_family="(CAGGA)n"  
repeat\_region  
22386..22530  
/rpt\_family="B1\_MM"  
repeat\_region  
22531..22556  
/rpt\_family="(CAGGA)n"

Query Match  
Best Local Similarity 100.0%; Score 18; DB 10; Length 190369;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTTCTACATATGCGCG 18

Db 159864 GTTCTACATATGCGCGC 159881

|||||

RESULT 5  
AKI09255  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) cDNA clone:002-185-A05, full insert sequence.  
677 bp mRNA linear PLN 24-JUL-2003

ACCESSION  
AKI09255  
VERSION  
AKI09255.1 GI:32994464  
KEYWORDS  
FUT\_CDNA; oligo capping.  
SOURCE  
Oryza sativa (japonica cultivar-group)  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
AUTHORS  
1  
The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team; Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yasaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yanagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group, Ootomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tanoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Marikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN; Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Kono, H., Miyazaki, A., Ootomo, Y., Oka, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.  
Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice  
Science 301 (5631), 376-379 (2003)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
2 (bases 1 to 677)  
12869764  
Aach, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, P., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Nariikawa, R., Nishikawa, J., Nishii, K., Nomura, K., Numaaki, R., Ohneda, E., Ohtsuki, K., Oka, M., Ooka, H., Ootomo, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Sato, K., Satoh, K., Satoh, K., Shibata, K., Shiraki, T., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yanagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yasaki, J., Yokomizo, S., and Yoshimura, A.  
Direct Submission  
Submitted (27-NOV-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp)  
Tel:81-29-838-7007, Fax:81-29-838-7007  
This clone is one of the 28k full-length cDNA clones from japonica rice.  
URL: [http://cdna01.dna.affrc.go.jp/cDNA/NIAS\\_Rice\\_Full-length\\_cDNA\\_Project\\_Team\\_Kikuchi,S.,Satoh,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yasaki,J.,](http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project_Team_Kikuchi,S.,Satoh,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yasaki,J.,)

Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yanagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., and Yamamoto, M.  
FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Nariikawa, R., Nishikawa, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tanoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.  
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, P., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numaaki, R., Ohtsuki, N., Ooka, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.  
Location/Qualifiers  
1. 677  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultiyar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="002-185-A05"

FEATURES  
source

ORIGIN  
Query Match 91.1%; Score 16.4; DB 8; Length 677;  
Best Local Similarity 94.4%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Cp 1 GTTCTACATATGCGCGC 18  
Db 329 GTTCTACATATGCGCGC 346  
|||||

RESULT 6  
AC014385/c  
LOCUS  
DEFINITION  
Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*.  
ACCESSION  
AC014385  
VERSION  
AC014385.1 GI:6436950  
KEYWORDS  
HTG; HTGS PHASE2.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 47636)  
Adams, M. and Venter, J.C.  
Direct Submission  
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
This sequence was identified as CDN:10210737 by the submitter.  
For further information on this sequence e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1. 47636  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"

FEATURES  
source

ORIGIN  
Query Match 91.1%; Score 16.4; DB 2; Length 47636;  
Best Local Similarity 94.4%; Pred. No. 2.1e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTTCTACATATGGCCG 18  
 |||||  
 DB 14262 GTTCTACTATATGGCCG 14245

RESULT 7  
 LOCUS DMBR1J17/c 92958 bp DNA linear HTG 11-OCT-1999  
 DEFINITION Drosophila melanogaster chromosome X clone BACR1J17 map 4A strain  
 Y; cn bw sp. \*\*\* SEQUENCING IN PROGRESS \*\*\*; 178 unordered pieces.  
 ACCESSION AL121811  
 VERSION AL121811.1 GI:6048214  
 KEYWORDS HTG: HTGS\_PHASE1.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Spheroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 92958)  
 RA Murphy, L., Harris, D. and Barrell, B.  
 JOURNAL Unpublished  
 REMARK Sequencing the X chromosome of Drosophila melanogaster  
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton Hall, Hinxton,  
 Cambridge CB10 1SA, U.K.  
 2 (bases 1 to 92958)  
 Benos, P.  
 REFERENCE Direct Submission  
 AUTHORS Submitted (10-OCT-1999) European Drosophila Genome Sequencing  
 JOURNAL Consortium  
 COMMENT This is a 'working draft' sequence. It currently  
 consists of 178 contigs. The true order of the pieces is not known  
 and their order in this sequence record is arbitrary. Gaps between  
 the contigs are represented as runs of N, but the exact sizes of  
 the gaps are unknown. This record will be updated with the finished  
 sequence. 1 415: contig of 415 in length  
 416 515: gap of unknown length  
 516 953: contig of 438 in length  
 954 1053: gap of unknown length  
 1054 1308: contig of 255 in length  
 1309 1408: gap of unknown length  
 1409 1448: contig of 40 in length  
 1449 1548: gap of unknown length  
 1549 2026: contig of 478 in length  
 2027 2126: gap of unknown length  
 2127 2326: contig of 200 in length  
 2327 2426: gap of unknown length  
 2427 2778: contig of 352 in length  
 2779 2878: gap of unknown length  
 2879 3125: contig of 247 in length  
 3126 3225: gap of unknown length  
 3226 3257: contig of 32 in length  
 3258 3357: gap of unknown length  
 3359 3720: contig of 363 in length  
 3721 3820: gap of unknown length  
 3821 4025: contig of 205 in length  
 4026 4125: gap of unknown length  
 4126 4575: contig of 450 in length  
 4576 4675: gap of unknown length  
 4676 4972: contig of 297 in length  
 4973 5072: gap of unknown length  
 5073 5504: contig of 432 in length  
 5505 5604: gap of unknown length  
 5606 5967: contig of 362 in length  
 5968 6067: gap of unknown length  
 6068 6401: contig of 335 in length  
 6402 6501: gap of unknown length  
 6502 6901: contig of 400 in length  
 6902 7001: gap of unknown length  
 7002 7024: contig of 23 in length  
 7025 7124: gap of unknown length  
 7125 7403: contig of 279 in length

7404 7503: gap of unknown length  
 7504 7721: contig of 218 in length  
 7722 7821: gap of unknown length  
 7823 7911: contig of 90 in length  
 7912 8011: gap of unknown length  
 8012 8472: contig of 461 in length  
 8473 8572: gap of unknown length  
 8573 8736: contig of 164 in length  
 8737 8836: gap of unknown length  
 8837 9144: contig of 308 in length  
 9145 9244: gap of unknown length  
 9245 9589: contig of 345 in length  
 9590 9689: gap of unknown length  
 9690 9764: contig of 75 in length  
 9765 9864: gap of unknown length  
 9865 10313: contig of 449 in length  
 10314 10413: gap of unknown length  
 10414 11309: contig of 895 in length  
 11309 11408: gap of unknown length  
 11409 11900: contig of 492 in length  
 11901 12000: gap of unknown length  
 12001 12308: contig of 308 in length  
 12309 12408: gap of unknown length  
 12409 12749: contig of 341 in length  
 12750 12850: gap of unknown length  
 12851 13308: contig of 459 in length  
 13309 13409: gap of unknown length  
 13410 13597: contig of 189 in length  
 13598 13697: gap of unknown length  
 13698 13963: gap of unknown length  
 13964 14063: contig of 266 in length  
 14064 14245: gap of unknown length  
 14246 14345: contig of 182 in length  
 14346 14490: gap of unknown length  
 14491 14590: contig of 145 in length  
 14591 14979: gap of unknown length  
 14980 15079: contig of 389 in length  
 15080 15473: gap of unknown length  
 15474 15573: contig of 394 in length  
 15574 15983: gap of unknown length  
 15984 16083: contig of 410 in length  
 16084 16107: gap of unknown length  
 16108 16207: contig of 24 in length  
 16208 16471: gap of unknown length  
 16472 16571: contig of 264 in length  
 16572 16973: gap of unknown length  
 16974 17073: contig of 402 in length  
 17074 17303: gap of unknown length  
 17304 17403: contig of 230 in length  
 17404 17756: gap of unknown length  
 17757 17856: contig of 353 in length  
 17857 18097: gap of unknown length  
 18098 18197: contig of 241 in length  
 18198 18647: gap of unknown length  
 18648 18748: contig of 450 in length  
 18749 19166: gap of unknown length  
 19167 19266: contig of 419 in length  
 19267 19399: gap of unknown length  
 19400 19499: contig of 133 in length  
 19500 19767: gap of unknown length  
 19768 19867: contig of 268 in length  
 19868 20202: gap of unknown length  
 20203 20302: contig of 335 in length  
 20303 20472: gap of unknown length  
 20473 20572: contig of 170 in length  
 20573 21017: gap of unknown length  
 21018 21117: contig of 445 in length  
 21118 21352: gap of unknown length  
 21353 21452: contig of 235 in length  
 21453 21668: gap of unknown length  
 21669 21868: contig of 316 in length  
 21869 21926: gap of unknown length  
 21927 22026: contig of 58 in length  
 22027 22026: gap of unknown length

22027 22323: contig of 297 in length  
 22324 22423: gap of unknown length  
 22424 22884: contig of 461 in length  
 22885 22984: gap of unknown length  
 22985 23168: contig of 184 in length  
 23169 23268: gap of unknown length  
 23269 23578: contig of 310 in length  
 23579 23678: gap of unknown length  
 23679 23987: contig of 309 in length  
 23988 24087: gap of unknown length  
 24088 24427: contig of 340 in length  
 24428 24527: gap of unknown length  
 24528 25057: contig of 530 in length  
 25058 25157: gap of unknown length  
 25158 25777: contig of 620 in length  
 25778 25877: gap of unknown length  
 25879 26450: contig of 553 in length  
 26431 26530: gap of unknown length  
 26531 26965: contig of 435 in length  
 26966 27065: gap of unknown length  
 27066 27853: contig of 788 in length  
 27854 27953: gap of unknown length  
 27954 28287: contig of 334 in length  
 28288 28387: gap of unknown length  
 28388 28852: contig of 465 in length  
 28853 28952: gap of unknown length  
 28953 29147: contig of 195 in length  
 29148 29247: gap of unknown length  
 29248 29585: contig of 338 in length  
 29586 29686: gap of unknown length  
 29686 29830: contig of 145 in length  
 29831 30137: contig of 207 in length  
 30138 30237: gap of unknown length  
 30238 30524: contig of 287 in length  
 30525 30624: gap of unknown length  
 30625 30933: contig of 309 in length  
 30934 31033: gap of unknown length  
 31034 31458: contig of 425 in length  
 31459 31558: gap of unknown length  
 31559 31902: contig of 344 in length  
 31903 32002: gap of unknown length  
 32003 32778: contig of 776 in length  
 32779 32878: gap of unknown length  
 32879 33449: contig of 571 in length  
 33450 33549: gap of unknown length  
 33550 34007: contig of 458 in length  
 34008 34107: gap of unknown length  
 34108 34871: contig of 764 in length  
 34872 34971: gap of unknown length  
 34972 35441: contig of 470 in length  
 35442 35541: gap of unknown length  
 35542 36137: contig of 596 in length  
 36138 36237: gap of unknown length  
 36238 36507: contig of 270 in length  
 36508 36607: gap of unknown length  
 36608 36821: contig of 214 in length  
 36822 36921: gap of unknown length  
 36922 37545: contig of 624 in length  
 37546 37645: gap of unknown length  
 37646 38331: contig of 686 in length  
 38332 38431: gap of unknown length  
 38432 38668: contig of 437 in length

Query Match 91.1%; Score 16.4; DB 2; Length 92958;  
 Best Local Similarity 94.4%; Pred. No. 2.2e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTTCTACTATATGGCGG 18  
 Db 73481 GTTCTACTATATGGCGG 73464

RESULT 8  
 AP003240  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

AP003240 144233 bp DNA linear PLN 31-AUG-2004  
 Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,  
 PAC clone: P0406G08.  
 AP003240 BA000010  
 AP003240.2 GI:17026065

Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Erihartoideae; Oryzaceae; Oryza.

1  
 Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,  
 Katayose, Y., Wu, J., Nishimura, Y., Cheng, Z., Nagamura, Y.,  
 Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,  
 Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,  
 Hamada, M., Harada, C., Hijioka, S., Honda, M., Ichikawa, Y.,  
 Iidouna, A., Iijima, M., Ikeda, M., Ikono, M., Ito, S., Ito, T., Ito, Y.,  
 Ito, Y., Iwabuchi, A., Kamiya, K., Katsura, W., Katagiri, S.,  
 Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maenaka, T.,  
 Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,  
 Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,  
 Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,  
 Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,  
 Yamagata, H., Yamane, H., Yoshiki, S., Yoshinaka, R., Yukawa, K.,  
 Zhong, H., Yama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,  
 Yano, M., Jiang, J. and Gojobori, T.  
 The genome sequence and structure of rice chromosome 1  
 Nature 420 (6913), 312-316 (2002)

22337376  
 12447438  
 2 (bases 1 to 144233)  
 Sasaki, T., Matsumoto, T. and Yamamoto, K.  
 Direct Submission  
 Submitted (19-FEB-2001) Takuji Sasaki, National Institute of  
 Agrobiological Sciences, Rice Genome Research Program, Kannondai  
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/,  
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
 On Nov 20, 2001 this sequence version replaced gi:13027270.  
 Genes were predicted from the integrated results of the following:  
 GENSCAN1.0, BLASTX2.0, BLASTX2.0 as well as SplicePredictor  
 (October 1998 version). The genomic sequence was searched against  
 NCBI Nonredundant Protein database, nr  
 (http://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at  
 RGP. Protein homologies of the coding regions were searched against  
 NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent  
 the identified cDNA sequences using BLASTN 2.0 with the  
 corresponding DBJ accession no. and RGP clone ID.  
 A gene with identity or significant homology to a protein is  
 classified based on the protein name to indicate the homology level  
 such as same name, 'putative-' and '-like protein'. A gene without  
 significant homology to any protein but with EST homology (covering  
 almost the entire length of partial sequence) is classified as an  
 'unknown' protein. A gene predicted with a gene prediction program  
 is classified as a 'hypothetical' protein.  
 The orientation of the sequence is from T7 to SP6 of the PAC clone.  
 This sequence of P0406G08 clone has an overlap with P0408C03 (DBJ:  
 AP003241) at 5' end and with P0446B05 (DBJ: AP003251) at 3' end.  
 Detailed information on overlap and assembly quality together with  
 annotation of this entry is available at  
 http://rpg.dna.affrc.go.jp/GenomeSeq.html.  
 Location/Qualifiers  
 1..144233  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="1"  
 /clone="P0406G08"  
 join(3501..3543,4068..4149,4324..4408)

FEATURES  
 source  
 gene









sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STS are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3369-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### FEATURES

##### source

```
1. 172784
   location/Qualifiers
   /organism="Drosophila melanogaster"
   /mol_type="genomic DNA"
   /db_xref="taxon:7227"
   /chromosome="X"
   /clone="RP98-45017"
```

#### ORIGIN

```
Query Match      91.1%; Score 16.4; DB 3; Length 172784;
Best Local Similarity 94.4%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 GTTCTACATATGCGCG 18
Db      169794 GTTCTACTATGCGCG 169811
```

RESULT 11  
AC107403/c  
LOCUS AC107403 181831 bp DNA linear INV 30-MAY-2002  
DEFINITION Drosophila melanogaster X BAC RP98-21M21 (Roswell Park Cancer Institute Drosophila BAC library) complete sequence.  
AC107403  
AC107403.6 GI:20986392  
HTG.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 181831)  
Muzny, D., Scherer, S., Adams, M.D., Holt, R.A., Evans, C.A., Gocayne, J.D., Tabor, P., Williamson, A., Homai, F.H., Dugan-Rocha, S.D., Sodergren, E.S., Hodgson, A.H., Chen, R.C., Ayalew, M., Scott, G.S., Worley, K.W., Amandlides, P.G., Bradow, R.C., Rogers, Y., An, H., Baldwin, D., Beeson, K.Y., Brown, M., Buhay, C., Buzam, D.A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K., Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E., Draper, H., Emery-Cohen, A., Ferriter, A., Gary, N.D.S., Houck, J., Hostin, D., Howland, T.J., Hume, J., Idegawa, C., Jatali, M., Kovar, C., Liu, W., Mattei, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B.,

Nelson, K.A., Ndasag, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V., Scheeler, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M., Xiang, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O., Wheeler, D., Weinstein, G., Gibbs, R. and Venter, J.C.  
Direct Submission

#### TITLE

##### JOURNAL

Unpublished  
2 (bases 1 to 181831)

##### REFERENCE

##### AUTHORS

Worley, K.C., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C., Alshrooke, S.L., Amarungu, H.C., Are, J.R., Banks, T., Barbarta, J., Benton, C., Bimuge, K., Blankenburg, K., Bonini, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carlton, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harrie, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Huiyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudas, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichteage, O., Lieu, C., Liu, J., Liu, W., Louiege, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., Meleod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S., Ogum, M., Okwunnu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Peters, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoochert, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vaquez, L., Vera, V., Villalob, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williams, A., Wlezyk, R., Wooden, S., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.  
Direct Submission  
Unpublished  
3 (bases 1 to 181831)  
REFERENCE  
Worley, K.C.  
Direct Submission  
Submitted (20-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 181831)  
REFERENCE  
Worley, K.C.  
Direct Submission  
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 181831)  
REFERENCE  
Worley, K.C.  
Direct Submission  
Submitted (20-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
6 (bases 1 to 181831)  
REFERENCE  
Worley, K.C.  
Direct Submission  
Submitted (30-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 20, 2002 this sequence version replaced gi:20514404.

#### COMMENT

INFORMATION: <http://www.bgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES  
source  
Location/Qualifiers

1..181831  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/chromosome="X"  
/clone="RP98-21M21"

misc\_feature  
complement(102930..181831)  
/note="Overlaps Bases 87026..165927 of clone AC023725"  
/function="clone overlap"

ORIGIN  
Query Match  
Best Local Similarity 91.1% Score 16.4; DB 3; Length 181831;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GTTCACATATATGGCCG 18  
134073 GTTCACTATATGCGCG 134056

RESULT 12  
AE003429/c  
LOCUS  
DEFINITION  
Drosophila melanogaster chromosome X, section 13 of 74 of the complete sequence.  
ACCESSION  
AE003429.3 GI:22831639  
KEYWORDS  
Drosophila melanogaster (fruit fly)  
SOURCE  
Drosophila melanogaster  
ORGANISM  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 309037)  
REFERENCE  
Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Mortan, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D., Man, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelsson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, C., Beasley, E.M., Beeson, K.Y., Benos, P.V., Bernan, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brotler, P., Burdick, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Dou, L.E., Evans, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Edwards, C.C., Ferraz, C., Ferrier, S., Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Iqbal, K.A., Kalish, B.E., Kodra, C.D., Kraft, C., Kennis, J.A., Ketchum, K.A., Kimmel, B.E., Kodra, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Nuzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nuskern, D.R., Pacle, J.M., Palazolo, M., Peltan, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Sanders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stadelton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, B., Wang, A.H., Wang, X., Wang, Z.Y., Wassenaar, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodgett, J., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, R.W., Rubin, G.M. and Venter, J.C.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
10731132  
2 (bases 1 to 309037)  
Celinker, S.E., Wheeler, D.A., Krommiller, B., Carlson, J.W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S.P., Fritze, E., Hodgson, A., George, R.A., Hoskins, R.A., Laverty, T., Murny, D.M., Nelson, C.R., Pacle, J.M., Park, S., Pfeiffer, B.D., Richards, S., Sodergren, E.J., Svirskas, R., Taber, P.E., Wan, K., Stapleton, M., Sutton, G.G., Venter, C., Weinstock, G., Scherer, S.E., Myers, E.W., Gibbs, R.A. and Rubin, G.M.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence  
Genome Biol. 3 (12), RESEARCH0079 (2002)  
22426065  
12537568  
3 (bases 1 to 309037)  
Wheeler, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S., Hradecky, P., Huang, Y., Kaminker, J.S., Millburn, G.H., Prochownik, S.E., Smith, C.D., Tupy, J.L., Whitfield, E.J., Bayraktaroglu, L., Bernan, B.P., Betencourt, B.R., Celinker, S.E., de Grey, A.D., Drysdale, R.A., Harris, N.L., Richter, J., Russo, S., Schroeder, A.J., Shu, S.Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W.M., Rubin, G.M. and Lewis, S.E.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review  
Genome Biol. 3 (12), RESEARCH0083 (2002)  
22426069  
12537572  
4 (bases 1 to 309037)  
Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Fritze, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., Ashburner, M. and Celinker, S.E.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
The transposable elements of the *Drosophila melanogaster* euchromatic genome: a genomics perspective  
Genome Biol. 3 (12), RESEARCH0084 (2002)  
22426070



Davila, M. I., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Demson, S., Detamco, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georgegeorgis, J., Geer, K., Gill, R., Grady, W., Guerin, W., Guetara, M., Gunaratne, P., Haland, M., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hudon, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huangk, S., Hume, J., Idolbid, D., Jackson, A., Jackson, L., Jacob, L., Jiang, X., Jiang, S., Johnson, B., Johnson, R., Jolyvet, A., Karapach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovat, C., Kowalski, C., Kraft, C. L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuehwa, L., Louisedge, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, B., Manjapa, P., Martin, K., Martin, R., Martinez, E., Mangun, B., Mayrhofer, S., McLeod, M. P., McNell, T. z., Meenen, R., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, K., Morris, S., Mundasa, M., Murphy, M., Naik, L., Nankervylle, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Pascarelli, C., Paul, H., Perez, A., Petrusagoun, A., Pal, S., Parks, K., Pasterkamp, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Quirao, M., Quiroz, Z., Rachlin, E., Reeves, K., Regier, M. A., Reish, R., Reilly, B., Reilly, W., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. U., Shetty, J., Shivatsbeyn, A., Slason, I., Sitter, C. D., Smajd, S., Shen, H., Sodergren, E., Song, X. Z., Steller, C., Soeta, V., Steele, M., Strong, R., Sutton, A., Swatek, A., Tabori, P., Taylor, C., Thomas, N., Thomas, S., Tinney, A., Trejos, Z., Umanit, K., Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Wlezey, R., Wooden, H., Wolley, K., Williams, G., Willison, R., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Zeddenhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,

Direct Submission

```

Consensus quality: 219055 bases at least Q30
Consensus quality: 220048 bases at least Q20
Estimated insert size: 231031; sum-of-coverage estimation
Quality coverage: 4x In Q20 bases; sum-of-coverage estimation
-----
NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 319026: contig of 319026 bp in length
* 319027 319126: gap of unknown length
* 319127 320168: contig of 1042 bp in length
* 320169 320268: gap of unknown length
* 320269 321355: contig of 1087 bp in length.
Location/Qualifiers
1..321355

```

```

misc_feature
/clone="CH230-166M22"
1..1190
/note="wgs_end_extension
clone_end:Sp6"
54939..55369
/note="clone_boundary
clone_end:Sp6
site:ECORI
end_sequence:BH301347"
/complement(279655..279925)
/note="clone_boundary
clone_end:T7
site:ECORI
end_sequence:BH301346"
285280..287738
/note="wgs_end_extension
clone_end:T7"
31757..319026
/note="wgs_end_extension
clone_end:T7"

```

Query Match	91.1%	Score 16.4	DB 2	Length 31355
Best Local Similarity	94.4%	Pred. No. 2.4e+02		
Matches 17	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Qy	1	GTTCTACATATGGCCCG	18	
Db	248822	GCTTACATATGGCCCG	248805	

RESULT 14	
AE016931	
LOCUS	
DEFINITION	AE016931 303750 bp DNA linear BCT 28-MAR-2003
ACCESSION	Bacteroides thetaiotaomicron VPI-5482, section 6 of 21 of the complete genome.
VERSION	AE016931 AE015928
KEYWORDS	AE016931.1 GI:29338516
SOURCE	.
ORGANISM	Bacteroides thetaiotaomicron VPI-5482 Bacteroides thetaiotaomicron VPI-5482 Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales; Bacteroidaceae; Bacteroides. 1 (bases 1 to 303750)
REFERENCE	Xu,J., Björnsell,M.K., Hinnrod,J., Deng,S., Carmichael,L.K., AUTHORS Chang,H.C., Hooper,L.V. and Gordon,G.I. TITLE A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis JOURNAL Science 299 (5615), 2074-2076 (2003)

PUBMED 12663928  
2 (bases 1 to 303750)  
XU J., BJURELLI M.K., HIMROD J., DENG S., CARNICHAEL L.K.,  
CHANG H.C., HOOPER L.V. and GORDON J.I.  
Direct Submission  
Submitted (31-Oct-2002) Department of Molecular Biology and  
Pharmacology, Washington University in St. Louis, 660 S. Euclid,  
St. Louis, MO 63110, USA

FEATURES  
source location/Qualifiers  
1..303750

gene /organism="Bacteroides thetaiotaomicron VPI-5482"  
/mol\_type="genomic DNA"  
/strain="VPI-5482"  
/db\_xref="taxon:226186"  
complement(87..1649)  
/locus\_tag="BT1210"  
complement(87..1649)  
/locus\_tag="BT1210"  
/codon\_start=1  
/transl\_table=11

CDS /product="cytochrome D ubiquinol oxidase subunit I"  
/protein\_id="AA076317.1"  
/db\_xref="GI:29338517"  
/translation="MIESIDTSLDMSRAQFAMTAMWIFVPLTGLAVMGIMETL  
YKTNSEFMKRTAKFMKLEGINFAIGATGLILEFGTMSYMFVGDIFGAPLA  
IEGLAFPMKATPIANMFGMDKVSKEPHLASTLTGIGATISAMWILVANMMHFV  
GMEFNPTVNEWDPMFVAVATSPVANKFPHVTSGWLGVPVVGISCMFLKKRNR  
EPALASIKIGAIRGLVSSSLAVWTGDSGYIATQPMKLAANGLYEGGTNVGLVI  
GMLNPEKTYNDKDPFLFRIEISMSFLAERVDVPEITNIIEGQYQDKQTTA  
LSAEKIERGTATGALAAVPAASAGHEBDAVAQVLOENMPYFGYGIKQNOVL  
PNVNLNFAFRIMVILGQYFLPFIVLFFYKKDLSKRMWMIALTITPLAYIACQ  
AGNVAVCGROPMAIRDMLEPTSAISKLVDGVSQATPFIFILFTWMLIAEIMVVA  
IKKGEPEISNR"

gene complement(1670..1906)  
/locus\_tag="BT1211"  
complement(1670..1906)  
/locus\_tag="BT1211"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA076318.1"  
/db\_xref="GI:29338518"  
/translation="MKNTLLAIWNPYLEGFRSMTLGRTLMLILVKLFIEMFIKIPF  
PPDLGHPMDADKGYVGNELIQRALPEKTTDF"

CDS 2132..3523  
/locus\_tag="BT1212"  
2132..3523  
/locus\_tag="BT1212"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane efflux protein"  
/protein\_id="AA076319.1"  
/db\_xref="GI:29338519"

gene /translation="MINVRLTVIAFLAGLLPGISAQESVQTAQSENNLPACOMDL  
OSCIDVALQONITIRKRVNAESSTOIVYTKAKALLPRLSPSTSONVNPYBESSIT  
ISGESEVLETSKTSYNGNVLGNASTYNGSKRLNTI KOEKLNNQANLNVESENSI  
QESIAQIYIQLIYAESVKVENTLQVSIARDGGLNAGSIKSDLAQLEQVST  
DRYQVLAQATLDQYKQLQLELDGENENNVLLPALSDEVNLLPLPTKQDVMSAL  
TLRPEIASKLSVASESLGIDIAAGVLPITISLAGIQTNTSGSDPTFGBOVNGN  
NSIGSVSVPIFNRRQTKSAVQAKLQRETSILSLDBOKTLVYTIIEGLMDANSAQO  
RYAANAEKLRSTQSYDLISBQFNLGMKNVTELLTEKNLLLAQOEQLQAKYVALMT  
OLKFEYQDKLAL"

gene 3575..4804  
/locus\_tag="BT1213"  
3575..4804  
/locus\_tag="BT1213"  
/codon\_start=1  
/transl\_table=11  
/product="periplasmic component of efflux system"  
/protein\_id="AA076320.1"  
/db\_xref="GI:29338520"

CDS /translation="MKTAKIILIAVAVVGGIWFPTGS PAKHVTMVASATVSKGD  
ISNSVATGTLIEPVIEVGTQVSGIIDIKIYVDYNSVYTKQGLIAEMDRVTLQSELAS

gene /organism="Bacteroides thetaiotaomicron VPI-5482"  
/mol\_type="genomic DNA"  
/strain="VPI-5482"  
/db\_xref="taxon:226186"  
complement(87..1649)  
/locus\_tag="BT1210"  
complement(87..1649)  
/locus\_tag="BT1210"  
/codon\_start=1  
/transl\_table=11

CDS /product="cytochrome D ubiquinol oxidase subunit I"  
/protein\_id="AA076317.1"  
/db\_xref="GI:29338517"  
/translation="MIESIDTSLDMSRAQFAMTAMWIFVPLTGLAVMGIMETL  
YKTNSEFMKRTAKFMKLEGINFAIGATGLILEFGTMSYMFVGDIFGAPLA  
IEGLAFPMKATPIANMFGMDKVSKEPHLASTLTGIGATISAMWILVANMMHFV  
GMEFNPTVNEWDPMFVAVATSPVANKFPHVTSGWLGVPVVGISCMFLKKRNR  
EPALASIKIGAIRGLVSSSLAVWTGDSGYIATQPMKLAANGLYEGGTNVGLVI  
GMLNPEKTYNDKDPFLFRIEISMSFLAERVDVPEITNIIEGQYQDKQTTA  
LSAEKIERGTATGALAAVPAASAGHEBDAVAQVLOENMPYFGYGIKQNOVL  
PNVNLNFAFRIMVILGQYFLPFIVLFFYKKDLSKRMWMIALTITPLAYIACQ  
AGNVAVCGROPMAIRDMLEPTSAISKLVDGVSQATPFIFILFTWMLIAEIMVVA  
IKKGEPEISNR"

gene complement(1670..1906)  
/locus\_tag="BT1211"  
complement(1670..1906)  
/locus\_tag="BT1211"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA076318.1"  
/db\_xref="GI:29338518"  
/translation="MKNTLLAIWNPYLEGFRSMTLGRTLMLILVKLFIEMFIKIPF  
PPDLGHPMDADKGYVGNELIQRALPEKTTDF"

CDS 2132..3523  
/locus\_tag="BT1212"  
2132..3523  
/locus\_tag="BT1212"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane efflux protein"  
/protein\_id="AA076319.1"  
/db\_xref="GI:29338519"

gene /translation="MINVRLTVIAFLAGLLPGISAQESVQTAQSENNLPACOMDL  
OSCIDVALQONITIRKRVNAESSTOIVYTKAKALLPRLSPSTSONVNPYBESSIT  
ISGESEVLETSKTSYNGNVLGNASTYNGSKRLNTI KOEKLNNQANLNVESENSI  
QESIAQIYIQLIYAESVKVENTLQVSIARDGGLNAGSIKSDLAQLEQVST  
DRYQVLAQATLDQYKQLQLELDGENENNVLLPALSDEVNLLPLPTKQDVMSAL  
TLRPEIASKLSVASESLGIDIAAGVLPITISLAGIQTNTSGSDPTFGBOVNGN  
NSIGSVSVPIFNRRQTKSAVQAKLQRETSILSLDBOKTLVYTIIEGLMDANSAQO  
RYAANAEKLRSTQSYDLISBQFNLGMKNVTELLTEKNLLLAQOEQLQAKYVALMT  
OLKFEYQDKLAL"

gene 3575..4804  
/locus\_tag="BT1213"  
3575..4804  
/locus\_tag="BT1213"  
/codon\_start=1  
/transl\_table=11  
/product="periplasmic component of efflux system"  
/protein\_id="AA076320.1"  
/db\_xref="GI:29338520"

CDS /translation="MKTAKIILIAVAVVGGIWFPTGS PAKHVTMVASATVSKGD  
ISNSVATGTLIEPVIEVGTQVSGIIDIKIYVDYNSVYTKQGLIAEMDRVTLQSELAS

gene /organism="Bacteroides thetaiotaomicron VPI-5482"  
/mol\_type="genomic DNA"  
/strain="VPI-5482"  
/db\_xref="taxon:226186"  
complement(87..1649)  
/locus\_tag="BT1210"  
complement(87..1649)  
/locus\_tag="BT1210"  
/codon\_start=1  
/transl\_table=11

CDS /product="putative ABC transporter, permease protein"  
/protein\_id="AA076322.1"  
/db\_xref="GI:29338522"  
/translation="MNGTNI FKIALRALANNKLRAPLTMGIIIGVASVITMLAIGQG  
SKSIIQOILESMNSNMIMHPGADRGRGVQDPSAMQTLTLTDEALRDETFSAIS  
PNVSSGGLQIEMSNNSVSVSGGLETLDRSLESQDMEFTEADIOSKAVCYIGT  
IYDNLPPDGSDDPIGKIIRFKIIPRVAVGQKAGVNSGODONIVIAVPTWKRLL  
AVTYLQGVAPASALSEDMTEBYATDEITRILRNHKKRASDDDPTRIQGLSTANST  
TDLMTLLACTAGISLVVGGIGIMNIMYVSTRKTRILKMSVGAAGVDILSGFLTE  
AIMISITGLIGVIGIGGASWIVSAVHWBIFIQPSVFLSFVACTVGVFCGVYPAK  
KAADLPDIEAIRYE"

gene 6883..7956  
/locus\_tag="BT1216"  
6883..7956  
/locus\_tag="BT1216"  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component system sensor protein,  
without kinase domain"

CDS /protein\_id="AA076323.1"  
/db\_xref="GI:29338523"  
/translation="MKQSLTSARRPLEALTHIIGWGMFGFPPEFVERENGINMAY  
LRHSAYLSPMIAFYVYFLVPRVLFQSQTKRYITYNILLCVIGMLHLMRSITPD  
PSFVPPKPRSGCGPGLFFVRDMLSVFTGLSAIRMSARMTQAEARKEABERNRSE  
AELKNRQNLNPHPLNTLNNIYALIAFOTDRAQAOVQELSKLRVLYXNOQYVPL  
GKETDFRNTYIELMRIRLSNVQMTTQIDILPDSTRILALIFITSLIENAFKGISFT  
ERSFPIHILAEINTEVYCEISNSNPNIMDKSGSGIGLEQVNRRLIILYPOGYTWOK  
GISEDGKEYSRISIRIESINK"

gene 8027..8725  
/locus\_tag="BT1217"  
8027..8725  
/locus\_tag="BT1217"  
/codon\_start=1  
/transl\_table=11  
/product="two-component system response regulator"

CDS /protein\_id="AA076324.1"  
/db\_xref="GI:29338524"  
/translation="WILRCAIVDDEPLAUGLLESYNNKTPFLQLTGKYSAAVQAMKEL  
PGSEVDLLFLDIQMPELNGLEFSQWDSRRIVFTTAFGQYALDGRVNALDYLKPI  
SYVDPLQANKALQWELVQKPEIDISI FVKSQDYKQVQELKIKIMYIEGLKDYIKYT  
EDDAKPIISLMSKMAEELPSSRFIRVRSFIVQDKIRVIDRGRIVFDKTYIPISD  
SYKOVFOAFIDERS"

gene 9147..9845  
/locus\_tag="BT1218"  
9147..9845  
/locus\_tag="BT1218"  
/codon\_start=1

CDS /translation="MKTAKIILIAVAVVGGIWFPTGS PAKHVTMVASATVSKGD  
ISNSVATGTLIEPVIEVGTQVSGIIDIKIYVDYNSVYTKQGLIAEMDRVTLQSELAS



```

/transl_table=11
/Product="conserved hypothetical protein, with conserved
domain"
/protein_id="AA076325.1"
/db_xref="GI:29338525"
/translation="MCFNHSMSAKAIKYAARVGRQSDVAEYOSILEDGYHNAAFTPE
RYPILTSSDEVQFNMGILPEFVSRSEDAATEIRMTLTNAADRIPEKPSFRPLNKA
CIYVSTDFEFMRHREGANKIPLYIYVKQEPISMGYIDRLMDGCEHEPSPSITTTT
TNSITDIIDTNRKPMALITQEEEEKWLNPSLSKAEISLNLKFTPEKDAIVINRDS
DKSSNDPPIYORA"
complement(10131..11081)
/locus_tag="BT1219"
complement(10131..11081)

```

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO76326.1"
/db_xref="GI:29338526"
/translation="MNPFQSCFGHIIPODVAGKQGENVFIYVNLITDSPTVDKVDY
CANFAMLRSMNRNFPDMQSCFTGFGALAMWRLIPDKKPELSTSEIKGEYTAIV
SPGDLPLHIAKQMGICFEPASITLBKIKAGVASVDETHRGVMDSKAIIKGVDSY
NPAVDENLHPAVIAGEGDADPAAGGSIVFOKRIHDMVAMNALPVEQDEKVIYGHKKND
VLSISEEKDGNAAHNAVITNIGDGLIVRAMPEANPANTSKECYGYFVIGASVFSTRML
EMMTFGSGAGVITDLDFSTAITGTLFVPSYDLDELGE"
/complement(11266..11955)
/locus_tag="BRI1220"
/complement(11266..11955)

```

```

/codon_start=1
/transl_table=1
/product="6-phosphogluconolactonase"
/protein_id="PA076327.1"
/db_xref="GI:29338527"
/translation="METARSLFHLVDIMNAEPKTEINVSGSGTPALMDIYANEY
ADITPWRKMLVWVERCVPEPDSDSNYGMRSLILGIVIPENVERIRGKVRKAKS
AVRSELSEVQSDYKNGKMEPDIYLLAGDDGHSSLPFQDELSSLDQIVTSTFPM
NGQRIIRLMGFPILITARIIVIPITGKARAVEVEICHSQDTPAIIAIAHNAELFEM
DAQNAVVRND"

```

Query	3	TTCTACATATGCGCGC	18		0	0	0
Best Local Similarity	100.0%	Freq. No.	4.2e+02				
Matches	16	Conservative	0				
		Mismatches	0				
		Indels	0				
		Gaps	0				
Score	16	DB	1				
Length	303750						
DB	107920	TTCTACATATGCGCGC	107935				

RESULT 15	AK063186				
LOCUS	AK063186				
DEFINITION	AK063186	549 bp	mRNA	linear	PLN 24-JUL-2003
ORIGIN	Oryza sativa (japonica cultivar-group)				
ACCESSION	AK063186				
VERSION	AK063186.1				
KEYWORDS	FLI cDNA; oligo-capping.				
SOURCE	Oryza sativa (japonica cultivar-group)				
ORGANISM	Oryza sativa (japonica cultivar-group)				
	Burayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
	Eleusineaceae; Oryzoideae; Oryzaeae; Oryza.				
REFERENCE	1				

1. The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yasaki, J., Ishikawa, M., Yamada, H., Ooka, H., Kojima, K., Nami, T., Ohneda, E., Yanao, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Otsuo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Naitikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikunaga, J.,

TITLE	JOURNAL	PUBMED	REFERENCE	AUTHORS
Collection, mapping, and annotation of over 28,000 cDNA clones from Japanese rice	Science	301 (5631), 376-379 (2003)		Ikedo, R., Iehibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusunegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Hara, J., Carninci, F., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Kikawa, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kigawa, I., Kondo, S., Konno, H., Miyazaki, A., Oasato, N., Oga, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.
2 (bases 1 to 549)				Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, F., Hayashi, K., Itoh, M., Kikawa, A., Kondo, S., Konno, H., Miyazaki, A., Oasato, N., Oga, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

Hori, F., Hotta, N., Iida, Y., Iida, Y., Iida, Y., Ikeda, R., Imamura, K., Imoto, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Konda, M., Koya, S., Kurihara, C., Kurosaki, I., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Miuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Naitakawa, R., Nikura, J., Nishi, K., Nomura, K., Nunasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Ose, T., Oka, Y., Ootomo, Y., Ryu, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shingawara, A., Shiraki, T., Shibiki, T., Sogade, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Toyota, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyota, T., Tannoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

Direct Submission  
 Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of  
 Agrobiological Sciences, Department of Molecular Genetics, Head of  
 Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki  
 305-8602, Japan (E-mail:skkikuchi@nias.affrc.go.jp,  
 Tel:81-29-838-7007, Fax:81-29-838-7007)  
 This clone is one of the 28K full-length cDNA clones from japonica  
 rice.

```

FEATURES
    source
        1. 549
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="mRNA"
            /cultivar="Nipponbare"

```



ORIGIN /db xref="taxon:39947"  
/clone="001-112-B04"

Query Match 85.6%; Score 15.4; DB 8; Length 549;  
Best Local Similarity 94.1%; Pred. No. 6.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTCTACATATGCGCCG 18  
|||||  
Db 311 TTCTACATCATGCGCCG 327

Search completed: March 2, 2005, 03:35:49  
Job time : 1826 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 1, 2005, 17:32:45 ; Search time 426 Seconds  
(without alignments)  
250.130 Million cell updates/sec

Title: US-10-677-982-1  
Perfect score: 18  
Sequence: 1 GTTCTACATATATGCGCG 18

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	18	100.0	18	AAx86621
2	15.4	85.6	1519	ABL16077
3	15.4	85.6	4927	ABL16076
4	15.4	85.6	110000	ABR03041.01
5	14.8	82.2	18	AAx86622
6	14.8	82.2	48	AAx86622
7	14.8	82.2	180	AAA94156
8	14.8	82.2	342	ADR60599
9	14.8	82.2	1302	ABD15713
10	14.8	82.2	1479	ABD15713
11	14.8	82.2	1566	ABD15757
12	14.8	82.2	1716	ABD15782
13	14.8	82.2	3057	ADSS0197
14	14.8	82.2	3057	ADSS5711
15	14.4	80.0	201	ADSS3863
16	14.4	80.0	201	ADSS3825
17	14.4	80.0	201	ADSS3721
18	14.4	80.0	201	ADSS3874
19	14.4	80.0	201	ADSS3791
20	14.4	80.0	201	ADSS3757

C	21	14.4	80.0	368	2	AAV20125
C	22	14.4	80.0	459	3	AAZ54318
C	23	14.4	80.0	1545	8	ACA50831
C	24	14.4	80.0	4076	13	ADSA9496
C	25	14.4	80.0	5174	5	AAST9419
C	26	14.4	80.0	6782	4	AAFT8309
C	27	14.4	80.0	6782	13	ADQ89799
C	28	14.4	80.0	7846	13	ADSS4820
C	29	14.4	80.0	7870	13	ADSS4824
C	30	14.4	80.0	7914	13	ADSS4821
C	31	14.4	80.0	7932	4	AAFT83910
C	32	14.4	80.0	7967	13	ADSS4822
C	33	14.4	80.0	8238	13	ADSS4819
C	34	14.4	80.0	34827	3	AAAB1481
C	35	14.4	80.0	67403	13	ADSS6460
C	36	14.4	80.0	97662	4	AAFT83908
C	37	14.4	80.0	110000	3	AAAB1489_3
C	38	14.4	80.0	110000	3	AAAB1489_4
C	39	14.4	80.0	110000	11	ACN43984_1
C	40	14.4	80.0	191284	12	ADQ97957
C	41	14.4	80.0	349980	3	AAFT21611
C	42	14.4	80.0	349980	3	AAFT21612
C	43	14	77.8	182	2	AAT20207
C	44	14	77.8	182	6	ABL67930
C	45	14	77.8	182	6	ABL63025

## ALIGNMENTS

RESULT 1	AAx86621	standard; DNA, 18 BP.
ID	AAx86621	
AC	AAx86621	
XX	15-OCT-1999	(first entry)
DT		
XX	Sonic hedgehog response element nucleotide sequence.	
DE		
XX	Sonic hedgehog response element; transcription factor; neuronal cell;	
XX	hedgehog-mediated signaling pathway; proliferation; differentiation;	
KW	proliferative disease; basal cell carcinoma; medulloblastoma; meningioma;	
KW	bone defect; polydactyly; jaw defect; rib defect; spina bifida;	
KW	familial midline defect; cyclopia; neural tube defect; ss.	
XX	Unidentified.	
OS		
XX	WO9941281-A1.	
PN		
XX	19-AUG-1999.	
PD		
XX	11-FEB-1999;	99WC-US003112.
PF		
XX	13-FEB-1998;	98US-00023249.
PR		
XX	(UWJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.	
PA	(BAYU ) BAYLOR COLLEGE MEDICINE.	
PA	Beachy PA, Tsai M, Tsai SY, Krishnan V, Chen C;	
PI	WPI; 1999-494517/41.	
XX		
DR		
XX		
PT	New transcription factors involved in hedgehog-mediated signaling, used	
PT	to identify modulators for controlling expression of target genes, e.g.	
PT	for treating cancer.	
XX		
PS	Claim 3; Page 49; 76pp; English.	
XX		
CC	The present sequence represents a sonic hedgehog response element. The	
CC	specification describes a transcription factor which is involved in a	
CC	hedgehog-mediated signaling pathway, has at least one phosphorylation	
CC	site and is phosphorylated in response to the pathway. This	

transcription factor binds to a hedgehog response element such as the  
present sequence. Modulation of the phosphorylation of the transcription  
factor is used to control expression of target genes involved in hedgehog  
-mediated signaling pathway, specifically for regulating proliferation or  
differentiation of neuronal cells; treatment of proliferative diseases  
(specifically basal cell carcinoma, medulloblastoma and meningioma); or  
to inhibit bone defects (e.g. polydactyly, jaw or rib defects, or spina  
bifida). Measuring the ratio of phosphorylated/dephosphorylated forms of  
transcription factor is used for diagnosis of hedgehog-mediated signaling  
pathway-mediated familial midline defects (specifically cyclopia or  
neural tube defects)

Sequence 18 BP; 4 A; 5 C; 4 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 18; DB 2; Length 18;  
Best Local Similarity 100.0%; Pred. No. 2,6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTTCACTAATGCGCG 18  
DB 1 GTTCACTAATGCGCG 18

## RESULT 2

ABLI6077/c  
ID ABLI6077 standard; cDNA; 1519 BP.

AC ABLI6077;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42713.

KW Drosophila; developmental biology; cell signalling; insecticide;

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR P-PSDB; ABB71974.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions.

XX Claim 1; SEQ ID NO 42713; 21bp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA  
sequences (ABLI1840-ABLI6175) and the encoded proteins (ABBS7737-  
ABBS72072). The sequence data for this patent did not form part of the  
printed specification, but was obtained in electronic format directly  
from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 1519 BP; 306 A; 393 C; 410 G; 410 T; 0 U; 0 Other;

Query Match 85.6%; Score 15.4; DB 4; Length 1519;  
Best Local Similarity 94.1%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TTCTACATTAATGCGCG 18  
DB 637 TTCTACATTAATGCGCG 621

## RESULT 3

ABLI6076/c  
ID ABLI6076 standard; cDNA; 4927 BP.

AC ABLI6076;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42710.

KW Drosophila; developmental biology; cell signalling; insecticide;

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR P-PSDB; ABB71973.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions.

XX Claim 1; SEQ ID NO 42710; 21bp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA  
sequences (ABLI1840-ABLI6175) and the encoded proteins (ABBS7737-  
ABBS72072). The sequence data for this patent did not form part of the  
printed specification, but was obtained in electronic format directly  
from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 4927 BP; 1288 A; 1106 C; 1081 G; 1452 T; 0 U; 0 Other;

Query Match 85.6%; Score 15.4; DB 4; Length 4927;  
Best Local Similarity 94.1%; Pred. No. 1.7e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TTCTACATTAATGCGCG 18  
DB 2620 TTCTACATTAATGCGCG 2604

## RESULT 4

ABA03041.01/c

Continuation (2 of 30) of ABA03041 from base 100001 (Listeria monocytogenes EGD-e genome  
WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041  
Fragment Name Begin End

```

WP ABA03041_00 1 110000
WP ABA03041_01 100001 210000
WP ABA03041_02 200001 310000
WP ABA03041_03 300001 410000
WP ABA03041_04 400001 510000
WP ABA03041_05 500001 610000
WP ABA03041_06 600001 710000
WP ABA03041_07 700001 810000
WP ABA03041_08 800001 910000
WP ABA03041_09 900001 1010000
WP ABA03041_10 1000001 1110000
WP ABA03041_11 1100001 1210000
WP ABA03041_12 1200001 1310000
WP ABA03041_13 1300001 1410000
WP ABA03041_14 1400001 1510000
WP ABA03041_15 1500001 1610000
WP ABA03041_16 1600001 1710000
WP ABA03041_17 1700001 1810000
WP ABA03041_18 1800001 1910000
WP ABA03041_19 1900001 2010000
WP ABA03041_20 2000001 2110000
WP ABA03041_21 2100001 2210000
WP ABA03041_22 2200001 2310000
WP ABA03041_23 2300001 2410000
WP ABA03041_24 2400001 2510000
WP ABA03041_25 2500001 2610000
WP ABA03041_26 2600001 2710000
WP ABA03041_27 2700001 2810000
WP ABA03041_28 2800001 2910000
WP ABA03041_29 2900001 2944528

```

```

Query Match 85.6%; Score 15.4; DB 6; Length 110000;
Best Local Similarity 94.1%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 TTCTACATATGCGCCG 18
Db 34251 TGCTACATATGCGCCG 34235

```

```

RESULT 5
AAK86622
ID AAK86622 standard; DNA; 18 BP.
XX
AC AAK86622;
XX
DT 15-OCT-1999 (first entry)
XX
DE Probe for wild type sonic hedgehog response element.
XX
KM Sonic hedgehog response element; transcription factor; neuronal cell;
KM hedgehog-mediated signaling pathway; proliferation; differentiation;
KM proliferative disease; basal cell carcinoma; medulloblastoma; meningioma;
KM bone defect; polydactyly; jaw defect; spina bifida;
KM familial midline defect; cyclopia; neural tube defect; probe; ss.
XX
OS Synthetic.
XX
PN W09941281-A1.
XX
PD 19-AUG-1999.
XX
PF 11-FEB-1999; 99WO-US003112.
XX
PR 13-FEB-1998; 98US-00023249.
XX
PA (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI Beechy PA, Tsai M, Tsai SY, Krishnan V, Chen C;
XX WPI; 1999-494517/41.
XX

```

```

PT New transcription factors involved in hedgehog-mediated signaling, used
PT to identify modulators for controlling expression of target genes, e.g.
PT for treating cancer.
XX
PS Example 8; Page 41; 76pp; English.
XX
CC The present sequence represents a probe for a wild type sonic hedgehog
CC response element. The specification describes a transcription factor
CC which is involved in a hedgehog-mediated signaling pathway, has at least
CC one phosphorylation site and is dephosphorylated in response to the
CC pathway. This transcription factor binds to a hedgehog response element
CC such as sequence AAK86621. Modulation of the phosphorylation of the
CC transcription factor is used to control expression of target genes
CC involved in hedgehog-mediated signaling pathway, specifically for
CC regulating proliferation or differentiation of neuronal cells; treatment
CC of proliferative diseases (specifically basal cell carcinoma,
CC medulloblastoma and meningioma); or to inhibit bone defects (e.g.
CC polydactyly, jaw or rib defects, or spina bifida). Measuring the ratio of
CC phosphorylated:dephosphorylated forms of transcription factor is used for
CC diagnosis of hedgehog-mediated signaling pathway-mediated familial
CC midline defects (specifically cyclopia or neural tube defects)
XX
SQ Sequence 18 BP; 2 A; 5 C; 6 G; 5 T; 0 U; 0 Other;

```

```

Query Match 82.2%; Score 14.8; DB 2; Length 18;
Best Local Similarity 88.9%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 GTTCTACATATGCGCCG 18
Db 1 GTTCTACATATGCGCCG 18

```

```

RESULT 6
AAA94156
ID AAA94156 standard; DNA; 48 BP.
XX
AC AAA94156;
XX
DT 30-JAN-2001 (first entry)
XX
DE Adenovirus 5 E1B promoter region vector PCR primer #1.
XX
KM Adenovirus 5; Ad 5; transcription factor Tcf binding site; colon cancer;
KM metastasis; liver tumour; colorectal cancer; gene therapy; PCR primer;
KM ss.
XX
OS Mastadenovirus.
XX
PN W0300056909-A1.
XX
PD 28-SEP-2000.
XX
PF 24-MAR-2000; 2000WO-GB001142.
XX
PR 24-MAR-1999; 99GB-00006815.
XX
PA (BTGT-) BTG INT LTD.
XX
PI Iggo R, Brunori M;
XX
PD WPI; 2000-628270/60.
XX
PT Viral DNA construct for treating neoplasms comprises tumor specific
PT transcription factor binding sites in place of wild type transcription
PT factor binding sites, operatively positioned in promoter region.
XX
PS Example; Page 30; 89pp; English.
XX
CC The present sequence is a PCR primer used during the construction of a
CC vector comprising the adenovirus 5 (Ad 5) E2 and E3 transcription sites.
CC This sequence was mutated so that the E1B promoter was replaced with 4
CC copies of the Tcf transcription factor binding site. It is an example of

```

CC the viral sequences of the invention, which are directed at tumour cells.  
 CC They are able to replicate so that they do not need to be given in large  
 CC quantities or inserted directly into the tumour, and those containing the  
 CC Tcf binding site are particularly useful in treating colon cancer. In  
 CC addition, metastases of the cancer, such as those found in the liver and  
 CC colorectal cancers can be treated using gene therapy in a similar way  
 XX

SO Sequence 48 BP; 9 A; 15 C; 7 G; 17 T; 0 U; 0 Other;

Query Match 82.2%; Score 14.8; DB 3; Length 48;  
 Best Local Similarity 88.9%; Pred. No. 2.1e+02;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTTCTACATATATGCGCG 18  
 DB 29 GTTCTATATATATGCGCG 46

RESULT 7  
 ID AAA94133 standard; DNA; 180 BP.  
 XX AAA94133;

AC AAA94133;

DT 30-JAN-2001 (first entry)

DE Adenovirus mutated E1B promoter.

KM Adenovirus transcription factor Tcf binding site; colon cancer;  
 XX metastasis; liver tumour; colorectal cancer; gene therapy; ds.

OS Mastadenovirus.

PN WO200056909-A1.

PD 28-SEP-2000.

PF 24-MAR-2000; 2000WO-GB001142.

PR 24-MAR-1999; 99GB-00006815.

PA (BTGI-) BTGI INT LTD.

PI Iggo R, Brunori M;

DR WPI; 2000-628270/60.

PT Viral DNA construct for treating neoplasms comprises tumor specific  
 PT transcription factor binding sites in place of wild type transcription  
 PT factor binding sites, operatively positioned in promoter region.  
 XX

PS Disclosure; Fig 21; 89pp; English.

CC The present sequence comprises the adenovirus E1B promoter. It has been  
 CC mutated to form one of the preferred sequences of the invention, which  
 CC are directed at tumour cells. These include sequences where the E2  
 CC promoter has been replaced with 4 copies of the Tcf transcription factor  
 CC binding site. They are able to replicate so that they do not need to be  
 CC given in large quantities or inserted directly into the tumour, and those  
 CC containing the Tcf binding site are particularly useful in treating colon  
 CC cancer. In addition, metastases of the cancer, such as those found in the  
 CC liver and colorectal cancers can be treated using gene therapy in a  
 CC similar way  
 XX

SO Sequence 180 BP; 42 A; 34 C; 40 G; 64 T; 0 U; 0 Other;

Query Match

Best Local Similarity 82.2%; Score 14.8; DB 3; Length 180;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTTCTACATATATGCGCG 18  
 DB 111 GTTCTATATATATGCGCG 128

RESULT 8  
 ID ADR60599 standard; cDNA; 342 BP.  
 XX ADR60599;

DT 02-DEC-2004 (first entry)

DE Cotton cDNA sequence, SEQ ID 1380.

KM Cotton; ss; plant; cold tolerance; growth rate; cell cycle pathway;  
 KM drought tolerance; plant disease resistance; galactomannan; lignin;  
 KM plant growth regulator; heat tolerance; herbicide tolerance;  
 KM homologous recombination; extreme osmotic condition tolerance;  
 KM pathogen resistance; pest resistance; yield; photosynthesis; seed oil;  
 KM stress resistance.  
 XX

OS Gossypium hirsutum.

PN US2004181830-A1.

PD 16-SEP-2004.

PF 29-JAN-2004; 2004US-00767795.

PR 07-MAY-2001; 2001US-00849529.

PR 12-DEC-2001; 2001US-00021323.

PA (KONA/) KOVALIC D K.

PA (ZHOU/) ZHOU Y.

PA (CAO/) CAO Y.

PI Kovalic DK, Zhou Y, Cao Y;

DR WPI; 2004-667718/65.

PT New recombinant nucleic acid molecules and polypeptides from Gossypium  
 PT hirsutum, useful for producing plants with improved biological  
 PT characteristics (e.g. improved plant cold or drought tolerance).  
 XX

PS Claim 1; SEQ ID NO 1380; 14pp; English.

CC The invention relates to a recombinant polynucleotide comprising any of  
 CC the 58798 Cotton plant cDNA sequences mentioned in the specification.  
 CC Also a recombinant polypeptide comprising any of the 58798 amino acid  
 CC sequences mentioned in the specification and producing a plant having an  
 CC improved property. Producing a plant having an improved property  
 CC comprises transforming a plant with a recombinant construct comprising a  
 CC promoter region functional in a plant cell operably joined to a  
 CC polynucleotide comprising a coding sequence for a polypeptide associated  
 CC with the property, and growing the transformed plant. The polypeptide is  
 CC useful for improving plant cold tolerance, manipulating growth rate in  
 CC plant cells by modification of the cell cycle pathway, improving plant  
 CC drought tolerance, providing increased resistance to plant disease,  
 CC producing galactomannan (or lignin or plant growth regulators), improving  
 CC plant heat tolerance, improving plant tolerance to herbicides, increasing  
 CC the rate of homologous recombination in plants, improving plant tolerance  
 CC to extreme osmotic conditions or to pathogens or pests, improving yield  
 CC by modification of photosynthesis, modifying seed oil or protein yield  
 CC and/or content, improving yield by modification of carbohydrate, nitrogen  
 CC or phosphorus use and/or uptake, or improving yield by providing improved  
 CC plant growth and development under at least one stress condition. The  
 CC polynucleotide and polypeptide may also be used in recombinant DNA  
 CC constructs, in physical arrays of molecules, as plant breeding markers,  
 CC or in computer-based storage and analysis systems. The present sequence  
 CC is a Cotton plant cDNA of the invention. NOTE: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from USPTO at  
 CC segdata.uspto.gov/sequence.html?DocID=20040181830. However only 6585  
 CC polynucleotide sequences were available, the remaining 52213  
 CC polynucleotides and all 58798 protein sequences were not present.

Query Match	82.2%	Score 14.8	DB 13	Length 342
Best Local Similarity	88.9%	Pred. No. 2.7e+02		
Matches 16	Conservative 0	Mismatches 2	Indels 0	Gaps 0

  

Sequence 342 BP, 97 A, 63 C, 66 G, 115 T, 0 U, 1 Other
<p>Query Match</p> <p>Best Local Similarity</p> <p>Matches 16</p> <p>Conservative 0</p> <p>Mismatches 2</p> <p>Indels 0</p> <p>Gaps 0</p>

  

Sequence 1302 BP, 238 A, 468 C, 386 G, 210 T, 0 U, 0 Other
<p>Query Match</p> <p>Best Local Similarity</p> <p>Matches 16</p> <p>Conservative 0</p> <p>Mismatches 2</p> <p>Indels 0</p> <p>Gaps 0</p>

[illegible]

AC ABD15757;  
XX  
XX 29-JUL-2004 (first entry)  
XX  
XX Pseudomonas aeruginosa polynucleotide #14361.  
DE  
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
XX US6551795-B1.  
XX  
XX 22-APR-2003.  
XX  
XX 18-FEB-1999; 99US-00252991.  
XX  
XX 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI: 2003-615309/58.  
DR P-PSDB; ABO82186.  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 14361; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biotech technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
XX  
SQ Sequence 1566 BP; 283 A; 523 C; 513 G; 247 T; 0 U; 0 Other;  
Query Match 82.2%; Score 14.8; DB 11; Length 1566;  
Best Local Similarity 88.9%; Pred. No. 3.2e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GTTCTACATATATGCGCG 18  
DB 112 GCTCTATATATATGCGCG 129  
RESULT 12  
ABD15782  
ID ABD15782 standard; DNA; 1716 BP.  
XX  
XX ABD15782;  
XX  
XX 29-JUL-2004 (first entry)  
XX  
XX Pseudomonas aeruginosa polynucleotide #14386.  
DE  
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX antibacterial.

XX  
XX Pseudomonas aeruginosa.  
OS  
XX US6551795-B1.  
XX  
XX 22-APR-2003.  
XX  
XX 18-FEB-1999; 99US-00252991.  
XX  
XX 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI: 2003-615309/58.  
DR P-PSDB; ABO82211.  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 14386; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biotech technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
XX  
SQ Sequence 1716 BP; 269 A; 499 C; 627 G; 321 T; 0 U; 0 Other;  
Query Match 82.2%; Score 14.8; DB 11; Length 1716;  
Best Local Similarity 88.9%; Pred. No. 3.3e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GTTCTACATATATGCGCG 18  
DB 1692 GCTCTATATATATGCGCG 1709  
RESULT 13  
ABD50197/c  
ID ABD50197 standard; CDNA; 3057 BP.  
XX  
XX ABD50197;  
XX  
XX 02-DEC-2004 (first entry)  
XX  
XX Bacterial polynucleotide #4940.  
DE  
XX Recombinant DNA construct; transformed plant; improved plant property;  
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;  
XX pathogen tolerance; pest tolerance; plant disease resistance;  
XX cell cycle pathway modification; plant growth regulator;  
XX homologous recombination; seed oil yield; protein yield; carbohydrate;  
XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
XX bacterial polynucleotide; gene; ss.  
OS  
XX Bacteria.



PN US2003233675-A1.  
 XX 18-DEC-2003.  
 XX 20-FEB-2003; 2003US-00369493.  
 XX 21-FEB-2002; 2002US-0360039P.  
 XX  
 PA (CAOY/) CAO Y.  
 PA (HINK/) HINKLE G J.  
 PA (SLAT/) SLATER S C.  
 PA (CHEN/) CHEN X.  
 PA (GOLD/) GOLDMAN B S.  
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 XX WPI; 2004-061375/06.  
 DR  
 XX  
 PT New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.  
 XX  
 PS Claim 1; SEQ ID NO 28627; 122pp; English.  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition. Improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polynucleotide used in  
 CC the scope of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification but was obtained in electronic  
 CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
 CC  
 XX  
 SQ Sequence 3057 BP; 652 A; 906 C; 949 G; 550 T; 0 U; 0 Other;  
 Query Match 82.2%; Score 14.8; DB 13; Length 3057;  
 Best Local Similarity 88.9%; Pred. No. 3.5e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTTCTACATATGGCGG 18  
 DB 2343 GTTGTACGAATGCGCG 2326  
 RESULT 14  
 ID AD555711/c  
 XX AD555711 standard; cDNA; 3057 BP.  
 AC AD555711;  
 XX  
 DT 02-DEC-2004 (first entry)  
 XX  
 DE Bacterial polynucleotide #7698.  
 XX  
 KM Recombinant DNA construct; transformed plant; improved plant property;  
 KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
 KM pathogen tolerance; pest tolerance; plant disease resistance;

KM cell cycle pathway modification; plant growth regulator;  
 KM homologous recombination; seed oil yield; protein yield; carbohydrate;  
 KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
 KM bacterial polynucleotide; gene; 88.  
 OS Bacteria.  
 XX  
 PN US2003233675-A1.  
 XX 18-DEC-2003.  
 XX 20-FEB-2003; 2003US-00369493.  
 XX 21-FEB-2002; 2002US-0360039P.  
 XX  
 PA (CAOY/) CAO Y.  
 PA (HINK/) HINKLE G J.  
 PA (SLAT/) SLATER S C.  
 PA (CHEN/) CHEN X.  
 PA (GOLD/) GOLDMAN B S.  
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 XX WPI; 2004-061375/06.  
 DR  
 XX  
 PT New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.  
 XX  
 PS Claim 1; SEQ ID NO 31385; 122pp; English.  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition. Improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polynucleotide used in  
 CC the scope of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification but was obtained in electronic  
 CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
 CC  
 XX  
 SQ Sequence 3057 BP; 652 A; 906 C; 949 G; 550 T; 0 U; 0 Other;  
 Query Match 82.2%; Score 14.8; DB 13; Length 3057;  
 Best Local Similarity 88.9%; Pred. No. 3.5e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTTCTACATATGGCGG 18  
 DB 2343 GTTGTACGAATGCGCG 2326  
 RESULT 15  
 ID AD538863/c  
 XX AD538863 standard; DNA; 201 BP.  
 AC AD538863;  
 XX

DT 16-DEC-2004 (first entry)  
XX  
DE Human autoimmune disease-related SNP context sequence - SEQ ID 4077.  
XX  
KW single nucleotide polymorphism detection; SNP detection;  
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;  
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;  
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;  
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;  
KW primary systemic vasculitis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2004083403-A2.  
XX  
PD 30-SEP-2004.  
XX  
PF 18-MAR-2004; 2004WO-US008461.  
XX  
PR 18-MAR-2003; 2003US-0455444P.  
XX 25-APR-2003; 2003US-0465241P.  
PA (APPL-) APPLERA CORP.  
XX  
PI Cargill M, Begovich AB, Alexander HC,  
XX  
XX WPI; 2004-728480/71.  
XX  
PT New isolated nucleic acid molecule comprises at least 8 contiguous  
PT nucleotides where one of the nucleotides is a single nucleotide  
PT polymorphism (SNP), useful for diagnosing or treating autoimmune  
PT diseases, e.g. rheumatoid arthritis.  
XX  
PS Claim 16; SEQ ID NO 4077; 123bp; English.  
XX  
CC The invention comprises amino acid and coding sequences containing  
CC genetic polymorphisms associated with an altered risk of developing an  
CC autoimmune disease (e.g. rheumatoid arthritis). The invention further  
CC comprises a method of identifying an individual that has an altered risk  
CC of developing an autoimmune disease, comprising detecting a single  
CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA  
CC and protein sequences of the invention are useful for diagnosing and  
CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1  
CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory  
CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious  
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,  
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The  
CC present DNA sequence represents a human autoimmune disease-related  
CC genomic-based SNP context sequence of the invention. NOTE: The present  
CC sequence is not shown in the specification, but has been retrieved from  
CC the WIPO website.  
XX  
SQ Sequence 201 BP; 52 A; 37 C; 46 G; 65 T; 0 U; 1 Other;

Query Match 80.0%; Score 14.4; DB 13; Length 201;  
Best Local Similarity 93.8%; Pred. No. 4.3e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTCTACATATGCGCC 17  
DB 125 TTCTACATATGCGCC 110

Search completed: March 2, 2005, 02:13:02  
Job time : 431 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 2, 2005, 02:07:02 ; Search time 128 Seconds  
(without alignments)  
230.101 Million cell updates/sec

Title: US-10-677-982-1  
Perfect score: 18  
Sequence: 1 GTTCTACATATGCGCCG 18

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/ina/5A COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	18	3	US-09-023-249-1
2	18	100.0	18	3	US-09-023-249-2
3	18	100.0	18	4	US-09-934-035-1
4	18	100.0	18	4	US-09-934-035-2
5	14.8	82.2	18	3	US-09-023-249-3
6	14.8	82.2	18	4	US-09-934-035-3
7	14.8	82.2	48	4	US-09-916-510A-30
8	14.8	82.2	180	4	US-09-916-510A-7
9	14.8	82.2	1302	4	US-09-252-991A-14317
10	14.8	82.2	1479	4	US-09-252-991A-14286
11	14.8	82.2	1565	4	US-09-252-991A-1461
12	14.8	82.2	1716	4	US-09-252-991A-1486
13	14.4	80.0	92387	4	US-09-949-016-1453
14	14.4	80.0	151295	4	US-09-949-016-14558
15	14.4	80.0	151295	4	US-09-949-016-14569
16	14.4	80.0	151295	4	US-09-949-016-14570
17	14.4	80.0	151295	4	US-09-949-016-14571
18	14.4	80.0	151295	4	US-09-949-016-14572
19	14.4	80.0	393753	4	US-09-949-016-14573
20	14.4	80.0	393753	4	US-09-949-016-14574
21	14.4	80.0	818128	4	US-09-949-016-14546
22	14.4	80.0	818128	4	US-09-949-016-14547
23	14.4	80.0	818128	4	US-09-949-016-14548
24	14.4	80.0	818128	4	US-09-949-016-14549
25	14.4	80.0	818128	4	US-09-949-016-14550
26	14.4	80.0	818128	4	US-09-949-016-14551
27	14.4	80.0	818128	4	US-09-949-016-14552

28	14.4	80.0	818128	4	US-09-949-016-14553	Sequence 14553, A
29	14.4	80.0	818128	4	US-09-949-016-14554	Sequence 14554, A
30	14.4	80.0	818128	4	US-09-949-016-14555	Sequence 14555, A
31	14.4	80.0	818128	4	US-09-949-016-14556	Sequence 14556, A
32	14.4	80.0	818128	4	US-09-949-016-14557	Sequence 14557, A
33	14.4	80.0	818128	4	US-09-949-016-14558	Sequence 14558, A
34	14.4	80.0	818128	4	US-09-949-016-14559	Sequence 14559, A
35	14.4	80.0	818128	4	US-09-949-016-14560	Sequence 14560, A
36	14.4	80.0	818128	4	US-09-949-016-14561	Sequence 14561, A
37	14.4	80.0	818128	4	US-09-949-016-14562	Sequence 14562, A
38	14.4	80.0	818128	4	US-09-949-016-14564	Sequence 14564, A
39	14.4	80.0	818128	4	US-09-949-016-14565	Sequence 14565, A
40	14.4	80.0	818128	4	US-09-949-016-14566	Sequence 14566, A
41	14.4	80.0	818128	4	US-09-949-016-14567	Sequence 14567, A
42	13.8	76.7	314	4	US-09-513-999C-13885	Sequence 33885, A
43	13.8	76.7	561	4	US-09-252-991A-12375	Sequence 12375, A
44	13.8	76.7	876	4	US-09-107-532A-551	Sequence 551, App
45	13.8	76.7	987	4	US-09-583-110-1890	Sequence 1890, App

ALIGNMENTS

RESULT 1  
US-09-023-249-1  
Sequence 1, Application US/09023249A  
Patent No. 6277566  
GENERAL INFORMATION:  
APPLICANT: Beachy, Philip A.,  
Tsal, Sophia Y.,  
Tsal, Ming-Jer,  
Krishnan, Venkatesh,  
Chen, Chien-Huan  
TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,249A  
FILING DATE: 13-Feb-1998  
CLASSIFICATION: Application  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/023,249  
FILING DATE: 13-Feb-98  
ATTORNEY/AGENT INFORMATION:  
NAME: Hallie, Lisa A.  
REGISTRATION NUMBER: 38,347  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 BASE PAIRS  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
FEATURE:  
NAME/KEY:  
LOCATION:  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-023-249-1

Query Match 100.0%; Score 18; DB 3; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTTCTACATTAATGCGCG 18  
Db 1 GTTCTACATTAATGCGCG 18

## RESULT 2

US-09-023-249-2/c  
; Sequence 2, Application US/09023249A  
; Patent No. 6273566  
; GENERAL INFORMATION:

APPLICANT: Beachy, Philip A.,  
Teal, Sophia Y.,  
Tsai, Ming-Jer,

Krishnan, Venkatesh,  
Chen, Chien-Huan  
TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:

ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92121

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,249A  
FILING DATE: 13-Feb-1998  
CLASSIFICATION: Application  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/023,249  
FILING DATE: 13-Feb-98  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: JHU1510-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 BASE PAIRS  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
FEATURE:  
NAME/KEY:  
LOCATION:

US-09-023-249-2  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Query Match 100.0%; Score 18; DB 3; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATTAATGCGCG 18  
Db 1 GTTCTACATTAATGCGCG 1

## RESULT 3

US-09-934-035-1  
; Sequence 1, Application US/09934035

; Patent No. 6733971  
; GENERAL INFORMATION:  
APPLICANT: Beachy, Philip A.,  
Teal, Sophia Y.,  
Tsai, Ming-Jer,  
Kishnan, Venkatesh,  
Chen, Chien-Huan  
TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/934,035  
FILING DATE: 21-Aug-2001  
CLASSIFICATION: Application  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/023,24949  
FILING DATE: 1998-02-13  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: JHU1510-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 BASE PAIRS  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
FEATURE:  
NAME/KEY:  
LOCATION:  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-934-035-1  
; Query Match 100.0%; Score 18; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTTCTACATTAATGCGCG 18  
Db 1 GTTCTACATTAATGCGCG 18

## RESULT 4

US-09-934-035-2/c  
; Sequence 2, Application US/09934035  
; Patent No. 6733971  
; GENERAL INFORMATION:

APPLICANT: Beachy, Philip A.,  
Teal, Sophia Y.,  
Tsai, Ming-Jer,

Krishnan, Venkatesh,  
Chen, Chien-Huan  
TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego

```

; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/934,035
; FILING DATE: 21-Aug-2001
; CLASSIFICATION: Application
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/023,24949
; FILING DATE: 1998-02-13
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: JH1510-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 BASE PAIRS
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY:
; LOCATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-934-035-2

Query Match      100.0%; Score 18; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTTCTACATAATGCGCCG 18
DB      18 GTTCTACATAATGCGCCG 1

RESULT 5
US-09-023-249-3
; Sequence 3, Application US/09023249A
; Patent No. 6277566
; GENERAL INFORMATION:
; APPLICANT: Beachy, Philip A.,
;          Tsai, Sophia Y.,
;          Tsai, Ming-Jer,
;          Krishnan, Venkatesh,
;          Chen, Chien-Huan
; TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,249A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: Application
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: 09/023,249
; FILING DATE: 13-Feb-98
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: JH1510-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 BASE PAIRS
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-023-249-3

Query Match      82.2%; Score 14.8; DB 3; Length 18;
Best Local Similarity 88.9%; Pred. No. 38;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTTCTACATAATGCGCCG 18
DB      1 GTTCTACGTGATGCGCCG 18

RESULT 6
US-09-934-035-3
; Sequence 3, Application US/09934035
; Patent No. 6733971
; GENERAL INFORMATION:
; APPLICANT: Beachy, Philip A.,
;          Tsai, Sophia Y.,
;          Tsai, Ming-Jer,
;          Krishnan, Venkatesh,
;          Chen, Chien-Huan
; TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/934,035
; FILING DATE: 21-Aug-2001
; CLASSIFICATION: Application
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/023,24949
; FILING DATE: 1998-02-13
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: JH1510-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 BASE PAIRS
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
```

SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-934-035-3

Query Match 82.2%; Score 14.8; DB 4; Length 18;  
Best Local Similarity 88.9%; Pred. No. 38;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTCTACATATATGCGCG 18  
DB 1 GTTCTACATATATGCGCG 18

RESULT 7  
US-09-916-510A-30  
Sequence 30, Application US/09916510A  
Patent No. 6544507  
GENERAL INFORMATION:  
APPLICANT: BRUNORI, MICHELE A.  
TITLE OF INVENTION: ANTI-NEOPLASTIC VIRAL AGENTS  
FILE REFERENCE: 604-596  
CURRENT APPLICATION NUMBER: US/09/916,510A  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: GB 9906815.7  
PRIOR FILING DATE: 1999-03-24  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 30  
LENGTH: 48  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Primer  
US-09-916-510A-30

Query Match 82.2%; Score 14.8; DB 4; Length 48;  
Best Local Similarity 88.9%; Pred. No. 46;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTCTACATATATGCGCG 18  
DB 29 GTTCTACATATATGCGCG 46

RESULT 8  
US-09-916-510A-7  
Sequence 7, Application US/09916510A  
Patent No. 6544507  
GENERAL INFORMATION:  
APPLICANT: IGGO, RICHARD D.  
TITLE OF INVENTION: ANTI-NEOPLASTIC VIRAL AGENTS  
FILE REFERENCE: 604-596  
CURRENT APPLICATION NUMBER: US/09/916,510A  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: GB 9906815.7  
PRIOR FILING DATE: 1999-03-24  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 7  
LENGTH: 180  
TYPE: DNA  
ORGANISM: Adenovirus VRS  
US-09-916-510A-7

Query Match 82.2%; Score 14.8; DB 4; Length 180;  
Best Local Similarity 88.9%; Pred. No. 58;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTCTACATATATGCGCG 18  
DB 111 GTTCTACATATATGCGCG 128

RESULT 9  
US-09-252-991A-14317/C  
Sequence 14317, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 14317  
LENGTH: 1302  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-14317

Query Match 82.2%; Score 14.8; DB 4; Length 1302;  
Best Local Similarity 88.9%; Pred. No. 83;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTCTACATATATGCGCG 18  
DB 139 GTTCTACATATATGCGCG 122

RESULT 10  
US-09-252-991A-14286/C  
Sequence 14286, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 14286  
LENGTH: 1479  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-14286

Query Match 82.2%; Score 14.8; DB 4; Length 1479;  
Best Local Similarity 88.9%; Pred. No. 85;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTCTACATATATGCGCG 18  
DB 158 GTTCTACATATATGCGCG 141

RESULT 11  
US-09-252-991A-14361  
Sequence 14361, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A

```
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 3142
; SEQ ID NO 14361
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14361
```

```
Query Match      82.2%; Score 14.8; DB 4; Length 1566;
Best Local Similarity 88.9%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 GTTCTACATATATGCGCG 18
Db      112 GCTCTATATATGCGCG 129
```

```
RESULT 12
US-09-252-991A-14386
; Sequence 14386, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 3142
; SEQ ID NO 14386
; LENGTH: 1716
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14386
```

```
Query Match      82.2%; Score 14.8; DB 4; Length 1716;
Best Local Similarity 88.9%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 GTTCTACATATGCGCG 18
Db      1692 GCTCTATATATGCGCG 1709
```

```
RESULT 13
US-09-949-016-14563
; Sequence 14563, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14,755
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14563
; LENGTH: 92387
```

```
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(92387)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14563
```

```
Query Match      80.0%; Score 14.4; DB 4; Length 92387;
Best Local Similarity 93.8%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 TTCTACATATGCGCC 17
Db      73070 TTTTACATATGCGCC 73085
```

```
RESULT 14
US-09-949-016-14568
; Sequence 14568, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14,755
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14568
; LENGTH: 151295
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(151295)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14568
```

```
Query Match      80.0%; Score 14.4; DB 4; Length 151295;
Best Local Similarity 93.8%; Pred. No. 3.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 TTCTACATATGCGCC 17
Db      73070 TTTTACATATGCGCC 73085
```

```
RESULT 15
US-09-949-016-14569
; Sequence 14569, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

```
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 14569
/ LENGTH: 151295
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(151295)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14569
```

## Query Match

80.0%; Score 14.4; DB 4; Length 151295;

Best Local Similarity 93.8%; Pred. No. 3.3e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 2 TTCTACATATGCGCC 17
    |||||
Db 73070 TTCTACATATGCGCC 73085
```

Search completed: March 2, 2005, 04:45:46  
Job time : 137 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 2, 2005, 02:13:11 ; Search time 4752 Seconds  
(without alignments)  
22.445 Million cell updates/sec

Title: US-10-677-982-1  
Perfect score: 18  
Sequence: 1 GTCTACATATGCGCCG 18

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5394803 seqs, 2962729879 residues

Total number of hits satisfying chosen parameters: 10789606

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*
- 20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	18	100.0	US-09-934-035-1	Sequence 1, Appli
2	18	100.0	US-09-934-035-2	Sequence 2, Appli
3	18	100.0	US-10-677-982-1	Sequence 1, Appli
4	18	100.0	US-10-677-982-2	Sequence 2, Appli
5	16.4	91.1	US-10-437-963-12834	Sequence 12834, A
6	16.4	91.1	US-10-437-963-12834	Sequence 12834, A
7	16.4	91.1	US-10-425-115-164709	Sequence 164709, A
8	15.4	85.6	US-10-425-115-42326	Sequence 42326, A
9	15.4	85.6	US-10-425-115-128320	Sequence 128320, A
10	15.4	85.6	US-10-424-599-4803	Sequence 4803, Ap
11	15.4	85.6	US-10-437-963-10575	Sequence 10575, A
			US-10-767-701-9687	Sequence 9687, Ap

12	15.4	85.6	1074	US-10-425-115-86671	Sequence 86671, A
13	14.8	82.2	18	US-09-934-035-3	Sequence 3, Appli
14	14.8	82.2	18	US-10-677-982-3	Sequence 3, Appli
15	14.8	82.2	48	US-09-916-510A-30	Sequence 30, Appli
16	14.8	82.2	48	US-10-376-630-30	Sequence 30, Appli
17	14.8	82.2	180	US-09-916-510A-7	Sequence 7, Appli
18	14.8	82.2	180	US-10-424-599-126750	Sequence 126750, A
19	14.8	82.2	315	US-10-424-599-126750	Sequence 1380, Ap
20	14.8	82.2	542	US-10-767-795-1380	Sequence 42225, A
21	14.8	82.2	599	US-10-437-963-42225	Sequence 74924, A
22	14.8	82.2	675	US-10-424-599-74924	Sequence 112918, A
23	14.8	82.2	710	US-10-425-115-112918	Sequence 28627, A
24	14.8	82.2	3057	US-10-369-493-28627	Sequence 31385, A
25	14.8	82.2	3682	US-10-369-493-31385	Sequence 32313, A
26	14.8	82.2	3741	US-10-425-115-112951	Sequence 112951, A
27	14.8	82.2	3741	US-10-425-115-112951	Sequence 61255, A
28	14.4	80.0	245	US-10-437-963-61255	Sequence 137088, A
29	14.4	80.0	468	US-10-424-599-137088	Sequence 16851, A
30	14.4	80.0	526	US-10-027-632-16851	Sequence 16851, A
31	14.4	80.0	526	US-10-027-632-16851	Sequence 197826, A
32	14.4	80.0	624	US-10-027-632-197826	Sequence 197826, A
33	14.4	80.0	624	US-10-027-632-197826	Sequence 23333, A
34	14.4	80.0	624	US-10-027-632-197826	Sequence 197827, A
35	14.4	80.0	624	US-10-027-632-197827	Sequence 38701, A
36	14.4	80.0	754	US-10-425-115-23333	Sequence 30829, A
37	14.4	80.0	1545	US-10-282-122A-38701	Sequence 34799, A
38	14.4	80.0	2163	US-10-425-115-34799	Sequence 27926, A
39	14.4	80.0	2360	US-10-369-493-27926	Sequence 205, App
40	14.4	80.0	4076	US-10-087-192-205	Sequence 8926, App
41	14.4	80.0	493631	US-10-674-124A-8926	Sequence 122, App
42	14	77.8	168	US-10-674-124A-8926	Sequence 295, App
43	14	77.8	182	US-09-954-531-295	Sequence 8843, Ap
44	14	77.8	182	US-10-425-115-8843	
45	14	77.8	301	US-10-425-115-8843	

#### ALIGNMENTS

RESULT 1  
US-09-934-035-1  
Sequence 1, Application US/09934035  
Patent No. US20020102646A1  
GENERAL INFORMATION:  
APPLICANT: Beachy, Philip A.,  
Tsai, Ming-Jer,  
Kriehnan, Venkatesh,  
Chen, Chien-Huan  
TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/934.035  
FILING DATE: 21-Aug-2001  
CLASSIFICATION: Application  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/023,24949  
FILING DATE: 1998-02-13  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347

```
REFERENCE/DOCKET NUMBER: JHU1510-1
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 858/677-1456
  TELEFAX: 858/677-1465
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 18 BASE PAIRS
      TYPE: nucleic acid
      STRANDEDNESS: both
      TOPOLOGY: linear
    MOLECULE TYPE: genomic DNA
    FEATURE:
      NAME/KEY:
      LOCATION:
      SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-934-035-1

Query Match
Best Local Similarity 100.0%; Score 18; DB 9; Length 18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTCTACATATATGCGCG 18
Db 1 GTTCTACATATATGCGCG 18

RESULT 2
US-09-934-035-2/c
; Sequence 2, Application US/09934035
; Patent No. US20020102646A1
GENERAL INFORMATION:
  APPLICANT: Beachy, Philip A.,
    Tsai, Sophia Y.,
    Krishnan, Venkatesh,
    Chen, Chien-Huan
  TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY
  NUMBER OF SEQUENCES: 8
  CORRESPONDENCE ADDRESSES:
    ADDRESSER: Gray Cary Ware & Freidenrich LLP
    STREET: 4365 Executive Drive, Suite 1600
    CITY: San Diego
    STATE: CA
    COUNTRY: U.S.A.
    ZIP: 92121
  COMPUTER READABLE FORM:
    MEDIUM TYPE: floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: Windows95
    SOFTWARE: Patent Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/934,035
    FILING DATE: 21-Aug-2001
    CLASSIFICATION: Application
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 09/023,24949
    FILING DATE: 1998-02-13
  ATTORNEY/AGENT INFORMATION:
    NAME: Halie, Lisa A.
    REGISTRATION NUMBER: 38,347
  REFERENCE/DOCKET NUMBER: JHU1510-1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 858/677-1456
    TELEFAX: 858/677-1465
  INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 18 BASE PAIRS
      TYPE: nucleic acid
      STRANDEDNESS: both
      TOPOLOGY: linear
    MOLECULE TYPE: genomic DNA
    FEATURE:
      NAME/KEY:
      LOCATION:
      SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-677-982-1

Query Match
Best Local Similarity 100.0%; Score 18; DB 9; Length 18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTCTACATATATGCGCG 18
Db 1 GTTCTACATATATGCGCG 18

RESULT 3
US-10-677-982-1
; Sequence 1, Application US/10677982
; Publication No. US20040082036A1
GENERAL INFORMATION:
  APPLICANT: Beachy, Philip A.,
    Tsai, Sophia Y.,
    Krishnan, Venkatesh,
    Chen, Chien-Huan
  TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY
  NUMBER OF SEQUENCES: 8
  CORRESPONDENCE ADDRESSES:
    ADDRESSER: Gray Cary Ware & Freidenrich LLP
    STREET: 4365 Executive Drive, Suite 1600
    CITY: San Diego
    STATE: CA
    COUNTRY: U.S.A.
    ZIP: 92121
  COMPUTER READABLE FORM:
    MEDIUM TYPE: floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: Windows95
    SOFTWARE: Patent Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/10/677,982
    FILING DATE: 01-Oct-2003
    CLASSIFICATION: Application
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US/09/023,249A
    FILING DATE: 13-Feb-1998
    APPLICATION NUMBER: 09/023,249
  ATTORNEY/AGENT INFORMATION:
    NAME: Halie, Lisa A.
    REGISTRATION NUMBER: 38,347
  REFERENCE/DOCKET NUMBER: JHU1510-1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 858/677-1456
    TELEFAX: 858/677-1465
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 18 BASE PAIRS
      TYPE: nucleic acid
      STRANDEDNESS: both
      TOPOLOGY: linear
    MOLECULE TYPE: genomic DNA
    FEATURE:
      NAME/KEY:
      LOCATION:
      SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-677-982-1

Query Match
Best Local Similarity 100.0%; Score 18; DB 17; Length 18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTCTACATATATGCGCG 18
Db 1 GTTCTACATATATGCGCG 18
```

RESULT 4  
US-10-677-982-2/c  
Sequence 2, Application US/10677982  
Publication No. US20040082036A1  
GENERAL INFORMATION:  
APPLICANT: Beachy, Philip A.,  
Teal, Sophia Y.,  
Teal, Ming-Jer,  
Krisman, Venkatesh,  
Chen, Chien-Huan  
TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/677,982  
FILING DATE: 01-Oct-2003  
CLASSIFICATION: Application  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,249A  
FILING DATE: 13-Feb-1998  
APPLICATION NUMBER: 09/023,249  
FILING DATE: 13-Feb-98  
ATTORNEY/AGENT INFORMATION:  
NAME: Halle, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: JH1510-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 BASE PAIRS  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
FEATURE:  
NAME/KEY:  
LOCATION:  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-677-982-2

Query Match 100.0%; Score 18; DB 17; Length 18;  
Best Local Similarity 100.0%; Pred. No. 5.5; 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
|||  
Db 18 GTTCTACATATGCGCG 1

RESULT 5  
US-10-437-963-12834  
Sequence 12834, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.,  
Kovalic, David K.,  
Zhou, Yihua  
APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 12834  
LENGTH: 231  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_18926C.1  
US-10-437-963-12834

Query Match 91.1%; Score 16.4; DB 18; Length 231;  
Best Local Similarity 94.4%; Pred. No. 60;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
|||  
Db 174 GTTCTACATATGCGCG 191

RESULT 6  
US-10-425-115-164709  
Sequence 164709, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.,  
Kovalic, David K.,  
Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 164709  
LENGTH: 767  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_81794C.1  
US-10-425-115-164709

Query Match 91.1%; Score 16.4; DB 18; Length 767;  
Best Local Similarity 94.4%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
|||  
Db 324 GTTCTACATATGCGCG 341

RESULT 7  
US-10-425-115-42326/c  
Sequence 42326, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.,  
Kovalic, David K.,  
Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28

```

? NUMBER OF SEQ ID NOS: 369326
? SEQ ID NO 42326
? LENGTH: 879
? TYPE: DNA
? ORGANISM: Zea mays
? FEATURE:
? OTHER INFORMATION: Clone ID: MRP4577_138602C.1
? US-10-425-115-42326

```

Query Match	91.1%	Score 16.4	DB 18	Length 879
Best Local Similarity	94.4%	Pred. No. 71		
Matches 17; Conservative	0	Mismatches 1	Indels 0	Gaps 0

```

Oy      1 GTTCTACATAATGCGCCG 18
          |||||
Db      583 GTTCTACATCATGCGCCG 566

```

```

RESULT 8
US-10-425-115-128320
/ Sequence 128320, Application US/10425115
/ Publication No. US20040214272A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yinhua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants
/ FILE REFERENCE: 38-21(53222)B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 128320
/ LENGTH: 379
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MRT4577_4849C.1
US-10-425-115-128320

```

	Query Match	Similarity	Score	DB	Length
Best Local	85.6%	94.1%	15.4	18	379
Matches	16	Conservative	0	Mismatches	1
				Indels	0
				Gaps	0

  

Qy	2	TTTCAATAAAGCGCG	18
Db	65	TTTCAATCAATGCGCG	81

```

RESULT 9
US-10-424-599-4803/c
; Sequence 4803, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 4803
; LENGTH: 533
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clome ID: PAT_MRT13847_104341C.1
US-10-424-599-4803

```

	85.6%	Score 15.4	DB 17	Length 533
Query March				
Best Local Similarity	94.1%	Pred. No. 2.4e+02		
Matches 16	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Oy	1	GTTCACATATACGCC	17	
Db	269	GTTCACATATACGCTCC	253	

RESULT 10  
US-10-437-963-10575  
; Sequence 10575, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION.

```

1  APPLICANT:  Dr. Ross, Thomas U.
2  APPLICANT:  Kovalic, David K.
3  APPLICANT:  Zhou, Yinnu
4  APPLICANT:  Cao, Yongwei
5  APPLICANT:  Wu, Wei
6  APPLICANT:  Boukharov, Andrey A.
7  APPLICANT:  Barbazuk, Brad
8  APPLICANT:  Li, Ping
9  TITLE OF INVENTION:  Rice Nucleic Acid Molecules and Other Molecules Associated with
10 TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
11 FILE REFERENCE:  38-21(53221) B
12 CURRENT APPLICATION NUMBER:  US/10/0437,963
13 CURRENT FILING DATE:  2003-05-14
14 NUMBER OF SEQ. ID NOS:  204966
15 SEQ ID NO 10575
16 LENGTH:  918
17 TYPE:  DNA
18 ORGANISM:  Oryza sativa
19 FEATURES:
20 OTHER INFORMATION:  Clone ID:  PAT_MRT4530_16883C.1
21 US-10-437-963-10575

```

	Query Match	85.6%	Score 15.4;	DB 18;	Length 918;
	Best Local Similarity	94.1%;	Pred. No. 2.6e+02;		
	Matches	16; Conservative	0; Mismatches	1;	Indels 0; Gaps 0;
Qy	2 TTCTCATATGACGGCCG	18			
Dd	332 TTCACATCATTGCGCG	348			

```

RESULT 11
US-10-767-701-9687
; Sequence 9687, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 9687
; LENGTH: 957
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS29420_1
US-10-767-701-9687

```

Query Match	85.6%	Score 15.4	DB 18	Length 957
Best Local Similarity	94.1%	Pred. No. 2.6e+02		
Matches 16	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	2 TTCTACATATAGCGCCG 18			

Db 357 TTCTACATCATGCCCG 373

## RESULT 12

US-10-425-115-86671  
Sequence 86671, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yinhua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 86671  
LENGTH: 1074  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_179051C.1  
US-10-425-115-86671

Query Match 85.6%; Score 15.4; DB 18; Length 1074;  
Best Local Similarity 94.1%; Pred. No. 2.6e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTCTACATCATGCCCG 18  
393 TTCTACATCATGCCCG 409

## RESULT 13

US-09-934-035-3  
Sequence 3, Application US/09934035  
Patent No. US2002010266A1  
GENERAL INFORMATION:  
APPLICANT: Beachy, Philip A.,  
Tsal, Sophia Y.,  
Tsal, Ming-Jer,  
Kriehnan, Venkatesh,  
Chen, Chien-Huan

TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/934,035  
FILING DATE: 21-Aug-2001  
CLASSIFICATION: Application

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/023,24949  
FILING DATE: 1998-02-13  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.

REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: JHU1510-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465

INFORMATION FOR SEQ ID NO: 3:

## SEQUENCE CHARACTERISTICS:

LENGTH: 18 BASE PAIRS

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: linear

MOLECULE TYPE: DNA

SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-934-035-3

Query Match 82.2%; Score 14.8; DB 9; Length 18;  
Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTTCTACATCATGCCCG 18  
Db 1 GTTCTACATCATGCCCG 18

## RESULT 14

US-10-677-982-3  
Sequence 3, Application US/10677982  
Publication No. US20040082036A1  
GENERAL INFORMATION:  
APPLICANT: Beachy, Philip A.,  
Tsal, Sophia Y.,  
Tsal, Ming-Jer,  
Kriehnan, Venkatesh,  
Chen, Chien-Huan

TITLE OF INVENTION: MODULATION OF HEDGEHOG-MEDIATED SIGNALING PATHWAY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/677,982  
FILING DATE: 01-Oct-2003  
CLASSIFICATION: Application

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,249A  
FILING DATE: 13-Feb-1998  
APPLICATION NUMBER: 09/023,249  
FILING DATE: 13-Feb-98  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.

REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: JHU1510-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 BASE PAIRS  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-10-677-982-3

Query Match 82.2%; Score 14.8; DB 17; Length 18;  
Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCCG 18  
 |||||  
 Db 1 GTTCTACGTATGCGCCG 18

## RESULT 15

US-09-916-510A-30  
 ; Sequence 30, Application US/0916510A  
 ; Patent No. US20020168349A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ICGO, RICHARD D.  
 ; TITLE OF INVENTION: ANTI-NEOPLASTIC VIRAL AGENTS  
 ; FILE REFERENCE: 604-596  
 ; CURRENT APPLICATION NUMBER: US/09/916,510A  
 ; CURRENT FILING DATE: 2001-07-30  
 ; PRIOR APPLICATION NUMBER: GB 9906815.7  
 ; PRIOR FILING DATE: 1999-03-24  
 ; NUMBER OF SEQ ID NOS: 48  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 30  
 ; LENGTH: 48  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Primer  
 US-09-916-510A-30

Query Match 82.2%; Score 14.8; DB 9; Length 48;  
 Best Local Similarity 88.9%; Pred. No. 3.9e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCCG 18  
 |||||  
 Db 29 GTCTATATATGCGCCG 46

Search completed: March 2, 2005, 06:05:04  
 Job time : 4755 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 1, 2005, 17:50:46 ; Search time 3122 Seconds  
(without alignments)  
219.461 Million cell updates/sec

Title: US-10-677-982-1  
Perfect score: 18  
Sequence: 1 GTTCTACATATGCGCG 18

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 60479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	18	100.0	497	7	CO223769
2	18	100.0	522	7	CF401440
3	18	100.0	530	7	CK443140
4	18	100.0	573	7	CF401389
5	18	100.0	608	7	CK443141
6	18	100.0	638	7	CO255135
7	18	100.0	674	7	CO209248
8	18	100.0	674	7	CO214192
9	18	100.0	727	7	BT648133
10	18	100.0	737	7	CF400196
11	18	100.0	807	7	CO213762
12	16.4	91.1	303	4	BG517706
13	16.4	91.1	478	7	CF400102
14	16.4	91.1	515	6	CA150538
15	16.4	91.1	524	2	BT121144
16	16.4	91.1	527	4	BT193194
17	16.4	91.1	585	8	CC156872
18	16.4	91.1	626	6	CA116627
19	16.4	91.1	626	6	CC653689
20	16.4	91.1	635	6	CD527173
21	16.4	91.1	666	6	CD484691
22	16.4	91.1	705	2	BF628429
23	16.4	91.1	718	9	CL192875
24	16.4	91.1	723	9	CL154062

25	16.4	91.1	724	9	CL192876
26	16.4	91.1	811	8	CC156871
27	16.4	91.1	826	9	CC658194
28	16.4	91.1	837	9	CG232932
29	16.4	91.1	828	9	CG232912
30	16.4	91.1	876	9	CG362961
31	16.4	91.1	1095	7	CG362970
32	16.4	91.1	1095	7	CK213111
33	16.4	91.1	1095	7	CK213111
34	15.4	85.6	254	2	AM607888
35	15.4	85.6	324	5	BU967700
36	15.4	85.6	345	8	BH098233
37	15.4	85.6	363	9	CG807442
38	15.4	85.6	390	7	CG531032
39	15.4	85.6	413	4	BG240324
40	15.4	85.6	421	6	BY676849
41	15.4	85.6	428	5	BX623930
42	15.4	85.6	500	1	AA528041
43	15.4	85.6	506	5	BX873595
44	15.4	85.6	507	4	BG240538
45	15.4	85.6	520	4	BG049487

## ALIGNMENTS

RESULT 1  
LOCUS CO223769/c 497 bp mRNA linear EST 22-JUN-2004  
DEFINITION WS01019.B21.015 SS-R-N-A-11 Picea sitchensis cDNA clone WS01019\_015  
3', mRNA sequence.

ACCESSION CO223769  
VERSION CO223769.1 GI:49046084  
KEYWORDS EST.

SOURCE Picea sitchensis (Sitka spruce)  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

REFERENCE 1. (bases 1 to 497)

AUTHORS Ralph, S., Kolesova, N., Cooper, D., Butterfield, Y., Kirkpatrick, R., Liu, J., Palmquist, D., Stott, J., Barber, S., Yang, G., Babak, R., Brown-John, M., Chand, S., Featherstone, R., Maeson, A., Mayo, M., Moran, J., Olson, T., Wong, D., Friedmann, M. F., Riland, C. E., Siddiqui, A., Holt, R., Jones, S., Marra, M., Ellis, B. E., Douglas, C., Riland, K. and Bohlmann, J.  
The spruce transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries  
Unpublished (2004)  
Contact: Joerg Bohlmann  
Genome BC forest genomics program  
University of British Columbia  
UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237,  
Vancouver, British Columbia, Canada, V6T 1Z3  
Tel: 1-604-822-0282  
Fax: 1-604-822-6097  
Email: bohlmann@interchange.ubc.ca  
Plate: WS01019 row: O column: 15  
High quality sequence stop: 497  
POLYAs.

## TITLE

JOURNAL COMMENT

FEATURES  
source

Location/Qualifiers  
1..497  
/organism="Picea sitchensis"  
/mol\_type="mRNA"  
/catalytic="GB2-229"  
/db\_xref="taxon:3332"  
/clone="WS01019\_015"  
/sex="Hermaphrodite"  
/tissue\_type="Young root growth (terminal 1-3 cm) and old root growth (distal to terminal 1-3 cm) tissues"  
/dev\_stage="three year old clonal trees grown under greenhouse conditions in standard potting soil mixture."  
/lab\_host="E. coli DH10B cells"  
/clone\_lib="SS-R-N-A-11"

/note="Organ: Roots; Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); RNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) by propagation. Normalization was applied according to published methods (Bonaldi M.F. et al. (1996) Genome Research 6(9):791) in order to reduce the abundance of highly expressed transcripts."

## ORIGIN

Query Match 100.0%; Score 18; DB 7; Length 497;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
|||||  
320 GTTCTACATATGCGCG 303

## RESULT 2

CF401440

## LOCUS

CF401440 522 bp mRNA linear EST 29-AUG-2003

## DEFINITION

RTW1\_12\_A07.91 A015 Well-watered loblolly pine roots WM1 Pinus

## ACCESSION

CF401440 taeda cDNA clone RTW1\_12\_A07\_A015 5', mRNA sequence.

## VERSION

CF401440.1 GI:34359857

## KEYWORDS

EST

## SOURCE

Pinus taeda (loblolly pine)

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

## REFERENCE

1 (bases 1 to 522) Pratt, L., Cordonnier-Pratt, M.-M., Lorenz, W.W., Dean, J.,

## AUTHORS

Gepermedhin, M., Dervins, C., Martin, T., White, T., Davis, J. and

## TITLE

An EST database from well-watered loblolly pine (Pinus taeda) roots

## JOURNAL

Unpublished (2003)

## COMMENT

Other ESTs: RTW1\_12\_A07.D1.A015

Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmprratt@uga.edu  
RNA prepared and library constructed by W. Walter Lorenz, School of Forestry, University of Georgia; plant material prepared at the University of Florida; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
Seq primer: JENREV (GAGGAACACGCTATACC).  
Location/Qualifiers

## FEATURES

source

1..522  
/organism="Pinus taeda"  
/mol\_type="mRNA"  
/strain="CCLONES"  
/db\_xref="taxon:3352"  
/clone="RTW1\_12\_A07\_A015"  
/lab\_host="DH10B-T1 Phage-resistant E. coli"  
/note="Vector: pSUI80; Site 1: EcoRI; Site 2: XhoI; The library was prepared from polyA+ RNA from loblolly pine (Pinus taeda) roots watered to pot capacity every other day. Pre-dawn water potential remained -0.3 Mpa +/-0.1."

## ORIGIN

Query Match 100.0%; Score 18; DB 7; Length 522;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
|||||  
422 GTTCTACATATGCGCG 439

## RESULT 3

CK443140

## LOCUS

CK443140 530 bp mRNA linear EST 08-JAN-2004

## DEFINITION

GQ0014b.BR.B01 GQ001: Male strobili developmental sequence Picea

## ACCESSION

CK443140 glauca cDNA clone GQ0014b\_B01 5', mRNA sequence.

## VERSION

CK443140.1 GI:40779514

## KEYWORDS

EST

## SOURCE

Picea glauca (white spruce)

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

## REFERENCE

1 (bases 1 to 530) Morency, M.-J., Cooke, J., Pavy, N., Parsons, L., Paule, C., Seguin, A.,

## AUTHORS

Retzel, E., Butterfield, Y., Barber, S., Yang, G., Stolt, J.,

## TITLE

Siddiqui, A., Holt, R., Mair, M. and Mackay, J.

## JOURNAL

Arborea Est sequencing in Picea glauca (white spruce)

## COMMENT

Unpublished (2004)

Contact: John Mackay  
Centre de Recherche en Biologie Forestiere  
Universite Laval  
Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4  
Fax: 418 656 7493  
Email: jmackay@rsvs.ulaval.ca  
Center for Computational Genomics and Bioinformatics (CCGB),  
University of Minnesota, MN id identifier: MN5159424 clone ID:  
GQ0014b\_B01 Clones available through: John Mackay, Ph. D.  
jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere  
(Forest Biology Research Center) Universite Laval Quebec, Quebec  
CANADA G1K 7P4  
Plate: 4b row: 01 column: B  
Seq primer: M13 Reverse Primer.  
Location/Qualifiers

## FEATURES

source

1..530  
/organism="Picea glauca"  
/mol\_type="mRNA"  
/strain="Tree 13-271"  
/db\_xref="taxon:3330"  
/clone="GQ0014b\_B01"  
/sex="Hermaphrodite"  
/tissue\_type="Entire strobilus"  
/dev\_stage="Three stages of preformed male strobili at end of winter dormancy were pooled: swollen fully closed buds, partly open buds and fully open buds"  
/lab\_host="E. coli DH10B cells"  
/clone\_id="GQ001: Male strobili developmental sequence"  
/note="Organ: Expanding male strobili; Vector: pBluescript II SK (+) XR; Site 1: Eco-R1; Site 2: Xho-I; cDNA was prepared from 5 mg of poly A+ selected RNA and was directionally ligated into the pBluescript II SK (+) XR vector (Stratagene), transformed by electroporation into DH10B cells (in vitro) for propagation"

## ORIGIN

Query Match 100.0%; Score 18; DB 7; Length 530;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 GTTCTACATATGCGCG 18  
 |||||  
 Db 317 GTTCTACATATGCGCG 334

RESULT 4  
 CF401389 573 bp mRNA linear EST 29-AUG-2003  
 LOCUS RTMW1\_12\_A07\_b1\_A015 Well-watered loblolly pine roots WM1 Pinus  
 DEFINITION taeda cDNA clone RTMW1\_12\_A07\_A015 3', mRNA sequence.

ACCESSION  
 VERSION CF401389.1 GI:34359806  
 KEYWORDS EST.  
 SOURCE Pinus taeda (loblolly pine)  
 ORGANISM Pinus taeda

REFERENCE  
 AUTHORS Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
 1 (bases 1 to 573)  
 Pratt, L., Cordonnier-Pratt, M.-M., Lorenz, W.W., Dean, J.,  
 Gebremedhin, M., Dervinis, C., Martin, T., White, T., Davis, J. and  
 Neale, D.

TITLE An EST database from well-watered loblolly pine (Pinus taeda) roots  
 JOURNAL Unpublished (2003)  
 COMMENT Other ESTs: RTMW1\_12\_A07\_g1\_A015  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu

FEATURES  
 source Location/Qualifiers  
 1..573  
 /organism="Pinus taeda"  
 /mol\_type="mRNA"  
 /strain="CCONES"  
 /db\_xref="taxon:3352"  
 /clone="RTMW1\_12\_A07\_A015"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Well-watered loblolly pine roots WM1"  
 /note="Vector: pSL1180; Site 1: EcoRI; Site 2: XhoI; The  
 library was prepared from polyA+ RNA from loblolly pine  
 (Pinus taeda) roots watered to pot capacity every other  
 day. Pre-dawn water potential remained -0.3 MPa +/-0.1.  
 Roots were harvested for RNA isolation. Double-stranded  
 cDNA was cloned unidirectionally into pSL1180. Inserts  
 excised with EcoRI (5' end) and XhoI (3' end)."

ORIGIN  
 Query Match 100.0%; Score 18; DB 7; Length 573;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
 |||||  
 Db 149 GTTCTACATATGCGCG 166

RESULT 5  
 CK443141 608 bp mRNA linear EST 08-JAN-2004  
 LOCUS GQ0014b\_TB\_B01 GQ001: Male strobili developmental sequence Picea  
 DEFINITION glauca cDNA clone GQ0014b\_B01 3', mRNA sequence.

ACCESSION  
 VERSION CK443141

VERSION CK443141.1 GI:40779516  
 EST.  
 KEYWORDS Picea glauca (white spruce)  
 SOURCE Picea glauca  
 ORGANISM Picea glauca

REFERENCE  
 AUTHORS Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.  
 1 (bases 1 to 608)  
 Morency, M.-J., Cooke, J., Pavy, N., Parsons, L., Paule, C., Seguin, A.,  
 Retzel, E., Butterfield, Y., Barber, S., Yang, G., Scott, J.,  
 Siddiqui, A., Holt, R., Marra, M. and Mackay, J.  
 Arborea EST sequencing in Picea glauca (white spruce)

TITLE Unpublished (2004)  
 JOURNAL Contact: John Mackay  
 COMMENT Centre de Recherche en Biologie Forestiere  
 Universite Laval  
 Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4  
 Fax: 418 656 7493  
 Email: jmackay@svs.ulaval.ca  
 Center for Computational Genomics and Bioinformatics (CCGB),  
 University of Minnesota, MN Id Identifier: MN5159425 Clone ID:  
 GQ0014b\_B01 Clones available through: John Mackay, Ph. D.  
 Professeur adjoint -Assistant professeur EMALI: Biologie Forestiere  
 (Forest Biology Research Center) Universite Laval Quebec, Quebec  
 CANADA G1K 7P4  
 Plate: 4b row: 01 column: B  
 Seq primer: PolyPlus Primer.

FEATURES  
 source Location/Qualifiers  
 1..608  
 /organism="Picea glauca"  
 /mol\_type="mRNA"  
 /strain="Tree 13-271"  
 /db\_xref="taxon:3330"  
 /clone="GQ0014b\_B01"  
 /sex="Hermaphrodite"  
 /tissue\_type="Entire strobilus"  
 /dev\_stage="Three stages of preformed male strobili at end  
 of winter dormancy were pooled: swollen fully closed buds,  
 partly open buds and fully open buds"  
 /lab\_host="E. coli DH10B cells"  
 /clone\_lib="GQ001: Male strobili developmental sequence"  
 /note="Organ: Expanding male strobili; Vector: pBluescript  
 II SK (+) XR; Site 1: Eco-RI; Site 2: Xho-I; cDNA was  
 prepared from 5 mg of poly A+ selected RNA and was  
 directionally ligated into the pluscript II SK (+) XR  
 vector (Stratagene), transformed by electroporation into  
 DH10B cells (In vitro) for propagation"

ORIGIN  
 Query Match 100.0%; Score 18; DB 7; Length 608;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
 |||||  
 Db 286 GTTCTACATATGCGCG 269

RESULT 6  
 CO255135 638 bp mRNA linear EST 23-JUN-2004  
 LOCUS WS00824\_B21\_E09 WS-X-N-A-9 Picea glauca cDNA clone WS00824\_E09 3',  
 DEFINITION mRNA sequence.

ACCESSION  
 VERSION CO255135  
 KEYWORDS CO255135.1 GI:49139136  
 SOURCE EST.  
 ORGANISM Picea glauca (white spruce)

REFERENCE  
 AUTHORS Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.  
 1 (bases 1 to 638)  
 Ralph, S., Kolosova, N., Cooper, D., Butterfield, Y., Kirkpatrick, R.,  
 Liu, J., Palmquist, D., Scott, J., Barber, S., Yang, G., Babakaliff, R.,

TITLE  
JOURNAL  
COMMENT  
BROWN-John, M., Chand, S., Featherstone, R., Masson, A., Mayo, M., Moran, J., Olson, T., Wong, D., Friedmann, M.F., Ritzland, C.E., Siddiqui, A., Holt, R., Jones, S., Marra, M., Ellis, B.E., Douglas, C., Ritzland, K. and Bohlmann, J.  
The spruce transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries  
Unpublished (2004)  
Contact: Joerg Bohlmann  
Genome BC forest genomics program  
University of British Columbia  
UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237, Vancouver, British Columbia, Canada, V6T 1Z3  
Tel: 1-604-822-0282  
Fax: 1-604-822-6097  
Email: bohlmann@interchange.ubc.ca  
Plate: WS00824 row: E column: 09  
High quality sequence step: 638  
POLYA=Yes

FEATURES  
source  
1. 638  
/location/Qualifiers  
/organism="Picea glauca"  
/mol\_type="mRNA"  
/cultivar="PG-29"  
/db\_xref="taxon:3330"  
/clone="WS00824\_E09"  
/sex="Hermaphrodite"  
/tissue\_type="Early season xylem harvested June 15th, mid season xylem harvested July 10th and late season xylem harvested August 17th"  
/lab\_host="E. coli DH10B cells"  
/clone\_1ib="WS-X-N-A-9"  
/note="Organ: Outer xylem from 25 year old trees harvested at Kalamalka Research Station in Vernon, British Columbia in 2001; Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

ORIGIN  
Query Match 100.0%; Score 18; DB 7; Length 638;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
|||||  
Db 327 GTTCTACATATGCGCG 310

RESULT 7  
CO209248/c 674 bp mRNA linear EST 21-JUN-2004  
LOCUS WS00914.B21 H17 IS-B-N-A-10 Picea engelmannii x Picea sitchensis  
DEFINITION cDNA clone WS00914\_H17 3', mRNA sequence.  
ACCESSION CO209248  
VERSION CO209248.1 GI:49021234  
KEYWORDS EST.  
SOURCE Picea engelmannii x Picea sitchensis  
ORGANISM Picea engelmannii x Picea sitchensis  
DEFINITION Bukaryoca; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Coniferales; Pinaceae; Picea.  
REFERENCE Ralph, S., Kolosova, N., Cooper, D., Butterfield, Y., Kirkpatrick, R., Liu, J., Palmquist, D., Scott, J., Barber, S., Yang, G., Babakoff, R.,

TITLE  
JOURNAL  
COMMENT  
BROWN-John, M., Chand, S., Featherstone, R., Masson, A., Mayo, M., Moran, J., Olson, T., Wong, D., Friedmann, M.F., Ritzland, C.E., Siddiqui, A., Holt, R., Jones, S., Marra, M., Ellis, B.E., Douglas, C., Ritzland, K. and Bohlmann, J.  
The spruce transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries  
Unpublished (2004)  
Contact: Joerg Bohlmann  
Genome BC forest genomics program  
University of British Columbia  
UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237, Vancouver, British Columbia, Canada, V6T 1Z3  
Tel: 1-604-822-0282  
Fax: 1-604-822-6097  
Email: bohlmann@interchange.ubc.ca  
Plate: WS00914 row: H column: 17  
High quality sequence step: 674  
POLYA=Yes

FEATURES  
source  
1. 674  
/location/Qualifiers  
/organism="Picea engelmannii x Picea sitchensis"  
/mol\_type="mRNA"  
/cultivar="Pal-1028"  
/db\_xref="taxon:273280"  
/clone="WS00914\_H17"  
/sex="Hermaphrodite"  
/lab\_host="E. coli DH10B cells"  
/clone\_1ib="IS-B-N-A-10"  
/note="Organ: Bark (with phloem and cambium attached) from one year old clonal trees grown under greenhouse conditions in standard potting soil mixture; Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); Bark was wounded using razor blades along the entire length of the tree at 5 mm intervals on opposite sides of the trunk. The same trees were also sprayed with a 0.01% (v/v) methyl jasmonate solution resuspended in 0.1% (v/v) tween 20 (50ms per tree). Bark tissue with phloem attached was harvested 3 hours, 6 hours, 12 hours, 24 hours, 2 days, 4 days and 8 days after initiating the treatment. Untreated control bark was also harvested at time 0 hours. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

ORIGIN  
Query Match 100.0%; Score 18; DB 7; Length 674;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
|||||  
Db 323 GTTCTACATATGCGCG 306

RESULT 8  
CO214192/c 674 bp mRNA linear EST 22-JUN-2004  
LOCUS WS00927.B21 D18 IS-B-N-A-10 Picea engelmannii x Picea sitchensis  
DEFINITION cDNA clone WS00927\_D18 3', mRNA sequence.  
ACCESSION CO214192  
VERSION CO214192.1 GI:49026638  
KEYWORDS EST.  
SOURCE Picea engelmannii x Picea sitchensis

ORGANISM *Picea engelmannii* x *Picea sitchensis*  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

REFERENCE  
 1 (bases 1 to 674)  
 Ralph, S., Kolosova, N., Cooper, D., Butterfield, Y., Kirkpatrick, R.,  
 Liu, J., Palomquist, D., Stott, V., Barber, S., Yang, G., Babakoff, R.,  
 Brown-John, M., Chand, S., Featherstone, R., Mason, A., Mayo, M.,  
 Moran, J., Olson, T., Wong, D., Friedmann, M.F., Ritland, C.E.,  
 Siddiqui, A., Holt, R., Jones, S., Marra, M., Ellis, B.E., Douglas, C.,  
 Ritland, K. and Bohlmann, J.  
 The spruce transcriptome: Analysis of expressed sequence tags from  
 multiple cDNA libraries  
 Unpublished (2004)

JOURNAL  
 COMMENT  
 Genome BC forest genomics program  
 University of British Columbia  
 UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237,  
 Vancouver, British Columbia, Canada, V6T 1Z3  
 Tel: 1-604-822-0282  
 Fax: 1-604-822-6097  
 Email: bohlmann@interchange.ubc.ca  
 Plate: WS00927 row: D column: 18  
 High quality sequence stop: 674  
 POLYAs=yes.

FEATURES  
 source  
 1..674  
 /organism="Picea engelmannii x Picea sitchensis"  
 /mol\_type="mRNA"  
 /culti\_var="Pal-1028"  
 /db\_xref="taxon:273280"  
 /clone="WS00927 D18"  
 /sex="Hermaphrodite"  
 /lab\_host="E. coli DH10B cells"  
 /clone\_lib="IS-B-N-A-10"  
 /note="Organ: Bark (with phloem and cambium attached) from  
 one year old clonal trees grown under greenhouse  
 conditions in standard potting soil mixture; Vector:  
 pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA);  
 Site 2: XhoI (3' end of cDNA); Bark was wounded using  
 razor blades along the entire length of the tree at 5 mm  
 intervals on opposite sides of the trunk. The same trees  
 were also sprayed with a 0.01% (v/v) methyl jasmonate  
 solution resuspended in 0.1% (v/v) tween 20 (~50mls per  
 tree). Bark tissue with phloem attached was harvested 3  
 hours, 6 hours, 12 hours, 24 hours, 2 days, 4 days and 8  
 days after initiating the treatment. Untreated control  
 bark was also harvested at time 0 hours. mRNA was isolated  
 from each tissue source independently and equal quantities  
 of mRNA from each tissue were then pooled. cDNA was  
 prepared from 5 micrograms of mRNA and directionally  
 ligated into the pBluescript II SK (+) XR vector using the  
 pBluescript II XR cDNA library construction kit according  
 to manufacturer's instructions with modifications  
 (Stratagene). Plasmid DNA was then transfected by  
 electroporation into DH10B cells (Invitrogen) for  
 propagation. Normalization was applied according to  
 published methods [Bonaldo M.P. et al. (1996) Genome  
 Research 6(9):791] in order to reduce the abundance of  
 highly expressed transcripts."

ORIGIN  
 Query Match 100.0%; Score 18; DB 7; Length 674;  
 Best Local Similarity 100.0%; Pred. NO. 25;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
 |||||  
 Db 323 GTTCTACATATGCGCG 306

RESULT 9  
 Bi649133/c  
 LOCUS Bi649133 727 bp mRNA linear EST 12-SEP-2001

DEFINITION 603278931F1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5319343 5',  
 mRNA sequence.

ACCESSION Bi649133  
 VERSION Bi649133.1 GI:15563369  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 1 (bases 1 to 727)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

JOURNAL  
 COMMENT  
 Contact: Robert Straubeberg, Ph.D.  
 Email: cgaaps-remail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LHAM1809 row: m column: 08  
 High quality sequence stop: 669.

FEATURES  
 source  
 1..727  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="129, C57BL/6J, FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5319343"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="10 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_Mam3"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Q1igo dt.  
 library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH  
 Reference for transgenic model: Xu et al., Nature Genetics  
 22, 37-43 (1999)."

ORIGIN  
 Query Match 100.0%; Score 18; DB 4; Length 727;  
 Best Local Similarity 100.0%; Pred. NO. 25;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCG 18  
 |||||  
 Db 531 GTTCTACATATGCGCG 514

RESULT 10  
 CF400196 737 bp mRNA linear EST 29-AUG-2003  
 LOCUS CF400196  
 DEFINITION taeeda cDNA clone RTMW1\_3\_G12\_A015 5', mRNA sequence.  
 ACCESSION CF400196  
 VERSION CF400196.1 GI:34358613  
 KEYWORDS EST.  
 SOURCE Pinus taeda (loblolly pine)  
 ORGANISM Pinus taeda

REFERENCE  
 1 (bases 1 to 737)  
 Pratt, L., Cordonnier-Pratt, M.-M., Lorenz, W.W., Dean, J.,  
 Gebremedhin, M., Dervinis, C., Martin, T., White, T., Davis, J. and  
 Neale, D.  
 An EST database from well-watered loblolly pine (*Pinus taeda*) roots  
 Unpublished (2003)  
 Other ESTs: RTMW1\_3\_G12\_b1\_A015  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

RNA prepared and library constructed by W. Walter Lorenz, School of Forestry, University of Georgia; plant material prepared at the University of Florida; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.  
Seq primer: JENREV (CAGGAACGCTATGACC).

#### FEATURES

source

```
1. .737
/organism="Pinus taeda"
/mol_type="mRNA"
/strain="CCLONES"
/db_xref="taxon:3352"
/clone="RTM1_3 G12 A015"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_1lb="Well-watered loblolly pine roots WM1"
/notes="Vector: pSL1180; Site 1: EcoRI; Site 2: XhoI; The library was prepared from polyA+ RNA from loblolly pine (Pinus taeda) roots watered to pot capacity every other day. Pre-dawn water potential remained -0.3 MPa +/-0.1. Roots were harvested for RNA isolation. Double-stranded cDNA was cloned unidirectionally into pSL1180. Inserts excised with EcoRI (5' end) and XhoI (3' end)."
```

#### ORIGIN

Query Match 100.0%; Score 18; DB 7; Length 737;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCCG 18  
|||||  
Db 363 GTTCTACATATGCGCCG 380

RESULT 11  
CO213762/c 807 bp mRNA linear EST 22-JUN-2004  
LOCUS  
DEFINITION  
WS00930.B21 A20 IS-B-N-A-10 Picea engelmannii x Picea sitchensis  
cDNA clone WS00930\_A20 3', mRNA sequence.

ACCESSION CO213762  
VERSION CO213762.1 GI:49026208  
SOURCE EST.  
ORGANISM Picea engelmannii x Picea sitchensis  
Picea engelmannii x Picea sitchensis  
Bukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

REFERENCE 1 (bases 1 to 807)  
AUTHORS Ralph,S., Kolojeva,N., Cooper,D., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stolt,J., Barber,S., Yang,G., Babakati,R., Brown-John,M., Chand,S., Featherstone,R., Maeson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Friedmann,M.F., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.F., Douglas,C., Ritland,K. and Bohlmann,J.

TITLE The spruce transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries  
JOURNAL COMMENT Unpublished (2004)  
CONTACT: Joerg Bohlmann  
Genome BC forest genomics program  
University of British Columbia  
UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237,  
Vancouver, British Columbia, Canada, V6T 1Z3

FEATURES  
source  
Tel: 1-604-822-0282  
Fax: 1-604-822-6097  
Email: bohlmann@interchange.ubc.ca  
Plate: WS00930 row: A column: 20  
High quality sequence stop: 807.  
Location/Qualifiers

source

```
1. .807
/organism="Picea engelmannii x Picea sitchensis"
/mol_type="mRNA"
/cultivar="Pal-1028"
/db_xref="taxon:273280"
/clone="WS00930_A20"
/sex="Hermaprodite"
/lab_host="E. coli DH10B cells"
/clone_1lb="IS-B-N-A-10"
/notes="Organ: Bark (with phloem and cambium attached) from one year old clonal trees grown under greenhouse conditions in standard potting soil mixture. Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); Bark was wounded using razor blades along the entire length of the tree at 5 mm intervals on opposite sides of the trunk. The same trees were also sprayed with a 0.01% (v/v) methyl jasmonate solution resuspended in 0.1% (v/v) tween 20 (~50mls per tree). Bark tissue with phloem attached was harvested 3 hours, 6 hours, 12 hours, 24 hours, 2 days, 4 days and 8 days after initiating the treatment. Untreated control bark was also harvested at time 0 hours. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."
```

#### ORIGIN

Query Match 100.0%; Score 18; DB 7; Length 807;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTACATATGCGCCG 18  
|||||  
Db 300 GTTCTACATATGCGCCG 283

RESULT 12  
BG517706 303 bp mRNA linear EST 30-MAR-2001  
LOCUS  
DEFINITION  
947070C11.Y1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA  
sequence.

ACCESSION BG517706  
VERSION BG517706.1 GI:13490942  
SOURCE EST.  
ORGANISM Zea mays  
Zea mays  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 303)  
AUTHORS Walbot V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford  
JOURNAL COMMENT Unpublished (1999)  
CONTACT: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA

FEATURES  
source  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 947070 row: C column: 11.  
Location/Qualifiers  
1. .303

```

/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/tissue_type="leaf and stem, including leaf base"
/dev_stage="2 week old seedling (3 leaves)"
/lab_host="Xtl-BLue"
/clone_idb="947 - 2 week shoot from Baykan lab"
/note="Organ: shoot; Vector: lambda ZAP (BlueScript SK-),
Site_1: EcoRI, Site_2: XhoI; Directionally cloned using
Stratagene's UniLap XR cDNA cloning kit with the 5' end
at the EcoRI site. The library represents 8 x 105
independent recombinant phage. The plants were greenhouse
grown."

```

## ORIGIN

Query Match	91.1%	Score 16.4	DB 4	Length 303
Best Local Similarity	94.4%	Pred. No. 2e+02		
Matches	17	Conservative	0	Mismatches 1
				Indels 0
				Gaps 0
QY	1 GTTCTACATATGGCCG	18		
Db	44 GTTCTACATATGGCCG	61		

## RESULT 13

LOCUS CF400102 478 bp mRNA linear EST 29-AUG-2003  
DEFINITION RTW1\_3 G12.b1.A015 Well1-watered Jobbolly pine roots W11 Pinus  
taeda cDNA clone RTW1\_3 G12.A015 3', mRNA sequence.

**ACCESSION**

VERSION CF400102.1 GI:34358519  
KEYWORDS EST.

## REFERENCES

ORGANISM

**CONCLUSION**

REFERENCE  
AUTHORS  
1 (pages 1 to 478)  
Pratt, J., Cordonnier-Pratt, M.-M., Lorenz, W.W., Dean, J.,  
Gebremedhin, M., Dervinis, C., Martin, T., White, T., Davis, J. and  
Neale, D.

## AUTHORS

JOURNAL	TITLE	COMMENT
unpublished (2003)	An EST database from well-watered loblolly pine ( <i>Pinus taeda</i> ) roots	
Other_ESTs: RTW01_3_G12_G1_A015		

**COMMENT**

Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmp@plant.uga.edu  
RNA prepared and library constructed by W. Walter Lorenz, School of  
Forestry, University of Georgia; plant material prepared at the  
University of Florida; sequencing done in the Laboratory for  
Genomics and Bioinformatics, University of Georgia. Sequence ends  
have been trimmed to exclude vector and regions below phred quality  
16. Three-prime sequences are presented as their reverse complement  
and have been trimmed to exclude polyA.  
Seq primer: M13-21 (TGTAAACGACGGCCAGT)  
POLYA=Yes.

## FEATURES

## Bouyce

```

1. .478
/organism="Pinus taeda"
/mol_type="mRNA"
/strain="CCLONES"
/db_xref="taxon:3352"
/clone="RMW1-G12_A015"
/lab_host="DH10B-G11 phage-resistant E. coli"
/clone_lib="Well-saturated loblolly pine roots WMI"
/note="Tector: psl1180; Site 1: EcorI; Site 2: XhoI. The library was prepared from poly(A+) RNA from loblolly pine (Pinus taeda) roots watered to pot capacity every other day. Pre-dawn water potential remained -0.3 MPa +/-0.1.

```

ORIGIN

Query Match	91.1%;	Score 16.4;	DB 7;	length 478;
Best Local Similarity	94.4%;	Pred. No. 2.1e+02;		
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

## QY

```

1 GTTCTACATAATGCCCG 18
  |||||
18 GTTTACATAATGCCCG 35

```

**Dib**

RESULT 14	CA150538	LOCUS	DEFINITION
	CA150538	515 bp	mRNA
	SCBFRR2019C07.g	R22 Saccharum officinarum	CDNA clone
	5', mRNA sequence.	SCBFRR2019C07	EST 24-SEP-2003

## ACCESSION

CA150538.1 GI:35053690

## KEYWORDS

**Saccharum officinarum**

## ORGANISM

REFERENCE  
1 (bases 1 to 515)  
complex.  
Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum  
complex.

## REFERENCES

## The libraries that made SUCEST

## JOURNAL

Contact: Arruda P

## CONCLUSIONS

Centro de Biologia Molecular e Engenharia Genética  
Universidade Estadual de Campinas  
Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
Tel: 55 19 3788 1137  
Fax: 55 19 3788 1089  
Email: parinda@unicamp.br  
Clone distribution: clone distribution information can be found  
through the Brazilian Clone Collection Center (BCCC) at  
<http://www.bcccenter.fcav.unesp.br>  
Plate: 019 row: C column: 07  
Seq primer: T7 Promoter Primer.

## FEATURES

1. .515

```

/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCBR22019C07"
/lab_host="DH10B"
/clone_lib="R22"
/name="Orogen: Shoot-root transition zone from young plants
(small insert library); Vector: pSPori; Site: 1: SalI;
Site: 2: NotI; An unidirectional cDNA library generated
from [Shoot-root transition zone from young plants (small
insert library)]. cDNA was prepared from poly(A+ mRNA
using Superscript Plasmid System Kit (Invitrogen). The
double-strand cDNAs were fractionated in a sepharose
CL-2B 40cm-columns and fragments sizing between 0.8 and
1.5 kb were directionally cloned into the vector. Details
of each source of RNA and library construction can be
obtained at http://succest.lad.ic.unimcap.br/public"

```

ORIGIN

Query Match	91.1%;	Score 16.4;	DB 6;	Length 515;
Best Local Similarity	94.4%;	Pred. No. 2.1e+02;		
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	GTTCCTACATATAGCGCGC	18	
Db	208	GTTCCTACATATAGCGCGC	225	

Search completed: March 2, 2005, 03:05:19  
 Job time : 3130 secs

RESULT 15  
 BE121144 524 bp mRNA linear EST 13-JUN-2000  
 LOCUS UI-R-CA0-bau-e-11-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone  
 DEFINITION  
 UI-R-CA0-bau-e-11-0-UI 3', mRNA sequence.  
 BE121144  
 BE121144.1 GI:8513249  
 EST.  
 Rattus norvegicus (Norway rat)  
 SOURCE  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE  
 AUTHORS Bernaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: Bento-soares@iowa.edu

The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a donatide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 normalized hippocampus library cDNA library preparation: M.B.  
 Soares Lab Clone distribution: clones will be available through  
 Research Genetics (www.resgen.com) The following repetitive  
 elements were found in this cDNA sequence: 180-208,  
 >(CAAA)n#Simple repeat  
 Seq primer: M13\_Foward  
 PolyA=yes.

#### FEATURES

source Location/Qualifiers  
 1..524  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-CA0-bau-e-11-0-UI"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="UI-R-CA0"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker: Site 1: Not I; Site 2: Eco RI; The UI-R-CA0  
 library is a subtracted library derived from the following  
 tissues: thalamus, cerebellum, hypothalamus, medulla,  
 pons, midbrain, cerebral cortex, corpus striatum, testis,  
 and hippocampus. For a detailed description of the  
 library from which this clone was derived, please visit  
 our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction  
 has been previously described in (Bernaldo, Lennon and  
 Soares, Genome Research 6:791-806, 1996)  
 TAG TISSUE=hippocampus  
 TAG LIB=UI-R-CA0  
 TAG\_SEQ=GATG

#### ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 524;  
 Best Local Similarity 94.4%; Pred. No. 2.1e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTTCTACATATGGCGC 18  
 | |||||  
 Db 380 GCTCTACATATGGCGC 397